Pharmacogenomic incidental findings in 308 families: The NIH Undiagnosed Diseases Program experience

Elizabeth M.J. Lee, BA^{1,2}, Karen Xu^{1,2}, Emma Mosbrook^{1,2}, Amanda Links, BS^{1,2}, Jessica Guzman, BA^{1,2}, David R. Adams, MD, PhD^{1,2}, Elise Flynn, BA^{1,2}, Elise Valkanas, BA^{1,2}, Camillo Toro, MD^{1,2}, Cynthia J. Tifft, MD, PhD^{1,2}, Cornelius F. Boerkoel, MD, PhD^{1,2}, William A. Gahl, MD, PhD^{1,2} and Murat Sincan, MD^{1,2}

Purpose: Using single-nucleotide polymorphism (SNP) chip and exome sequence data from individuals participating in the National Institutes of Health (NIH) Undiagnosed Diseases Program (UDP), we evaluated the number and therapeutic informativeness of incidental pharmacogenetic variants.

Methods: Pharmacogenomics Knowledgebase (PharmGKB) annotated sequence variants were identified in 1,101 individuals. Medication records of participants were used to identify individuals prescribed medications with a genetic variant that might alter efficacy.

Results: A total of 395 sequence variants, including 19 PharmGKB 1A and 1B variants, were identified in SNP chip sequence data, and 388 variants, including 21 PharmGKB 1A and 1B variants,

were identified in the exome sequence data. Nine participants had incidental pharmacogenetic variants associated with altered efficacy of a prescribed medication.

Conclusions: Despite the small size of the NIH UDP patient cohort, we identified pharmacogenetic incidental findings potentially useful for guiding therapy. Consequently, groups conducting clinical genomic studies might consider reporting of pharmacogenetic incidental findings.

Genet Med advance online publication 2 June 2016

Key Words: next generation sequencing; NIH undiagnosed diseases program; pharmacogenomics; precision medicine; secondary findings

Incidental or secondary genetic findings are variants with medical or social implications discovered during genetic testing for an unrelated indication.¹ Recent discussions and a report by the American College of Medical Genetics and Genomics (ACMG) Working Group on Incidental Findings in Clinical Exome and Genome Sequencing have focused on disease-associated genes but not genetic determinants of drug metabolism.² Given that genomic variation influences human responsiveness to many drugs and contributes to phenotypes ranging from life-threatening adverse drug reactions to lack of therapeutic efficacy,³ the return of pharmacogenetic incidental findings has potentially significant medical benefit.⁴

The Clinical Pharmacogenetics Implementation Consortium (CPIC), which develops guidelines for incorporating pharmacogenomics findings into clinical practice, has identified variant–drug associations of high concern for clinicians and provides drug-dosing guidelines based on patient genotype. Therefore, we hypothesized that designing and implementing a process to identify pharmacological incidental findings in the genomic data generated by the National Institutes of Health (NIH) Undiagnosed Diseases Program⁵ (UDP) provide information to the medical community regarding the quantity and quality of pharmacogenetic incidental findings.

The NIH UDP routinely conducts single-nucleotide polymorphism (SNP) chip and exome sequencing analyses on probands and their family members. These data can be analyzed

for pharmacogenetic incidental findings based on variant–drug associations listed in the Pharmacogenomics Knowledgebase (PharmGKB). The typical number of such pharmacogenetic incidental findings has not been widely studied, particularly when family members other than the proband are included in diagnostic studies. Consequently, more data are needed to assess the possible impact and need for resources.

To delineate the impact of identifying pharmacogenetic incidental findings, we analyzed SNP chip data from 1,101 individuals derived from 308 families and research exome sequence data from 645 individuals derived from 158 families. For the 868 pharmacogenetic loci listed in the PharmGKB, we identified 949 independent pharmacogenetic findings using the SNP chip and exome sequence data and found that each individual had at least one. For nine individuals, these constituted incidental findings relevant to a medication that they were using. These data refine strategies for reporting of pharmacogenetic incidental findings.

MATERIALS AND METHODS

Subject cohort

Family members gave informed consent or assent under protocol 76-HG-0238, “Diagnosis and Treatment of Patients with Inborn Errors of Metabolism and Other Genetic Disorders,” approved by the National Human Genome Research Institute (NHGRI) institutional review board. The SNP chip data

¹NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, Maryland, USA; ²National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA. Correspondence: Murat Sincan (sincanm@mail.nih.gov)

Submitted 13 November 2015; accepted 8 March 2016; advance online publication 2 June 2016. doi:10.1038/gim.2016.47

were derived from a cohort of 308 families consisting of 355 affected individuals, 278 unaffected siblings, 459 unaffected parents, and 9 other unaffected family members; this cohort included 564 females and 537 males. The exome sequence data were from a subset of this cohort; 158 families contained 182 affected individuals, 150 unaffected siblings, 313 unaffected parents, and 326 females and 319 males. The average and median ages of the subjects at time of SNP chip analysis were 36.1 (SD 22.4) and 36.0 years, respectively. The average and median ages at the time of exome analysis were 35.2 (SD 22.4) and 36.0 years, respectively. Some subjects were deceased at the time of study; for those subjects, projected age at time of sequencing was used, since it is anticipated that incidental findings will be sought only in living subjects. Self-reported ancestry was white/European (75.3%), black/African American (2.5%), American Indian or Alaskan Native (0.4%), Asian (2.1%), multiracial (5.3%), and unknown (14.4%). These families included all those admitted to the NIH Undiagnosed Diseases Program and who had SNP chip or exome analyses. The SNP chip genotyping and exome sequencing were performed on a research basis between 2009 and 2014, not in a Clinical Laboratory Improvement Amendments–certified fashion.

DNA extraction

Genomic DNA was extracted from patients’ peripheral whole blood using the Gentra Puregene Blood kit (Qiagen, Valencia, CA).

SNP chip analysis

The extracted DNA was submitted to the NHGRI core lab and run on the Human OmniExpressExome v1.2 SNP oligo array (Illumina, San Diego, CA). Each family was analyzed individually following the UDP’s standard operating procedure. Call rates were >98% before quality control processing and were typically >99.7%.

Exome sequencing

Genomic DNA was submitted for exome sequencing using the Illumina TruSeq exome capture kit (Illumina), which targets roughly 60 million bases consisting of the Consensus Coding Sequence annotated gene set as well as some structural RNAs. Captured DNA was sequenced on the Illumina HiSeq platform until coverage was sufficient to call high-quality genotypes at 85% or more of targeted bases.

Alignment and genotype calling

Beagle software version 4 and the 1000 Genome Project’s HapMap data were used to generate a phased and imputed variant call format (VCF) file from SNP chip data for the parents and offspring.⁶ The VCF file was then used by AlleleSeq version 0.2.3 (ref. 7) to modify the human reference and create a maternal reference and a paternal reference, which are concatenated to generate a parental reference. Patient short reads from exome sequencing were aligned to all three reference sequences with Novoalign version 2.08.03 and lifted back over to the standard human reference using custom Java code. BAM files were

Figure 1 Flow chart summarizing the NIH Undiagnosed Diseases Program analysis of and observations for the pharmacogenetic variants listed in PharmGKB. The observations were derived from analysis of SNP chip and exome sequence data. The SNP chip data were derived from a cohort of 308 families consisting of 355 affected individuals, 278 unaffected siblings, 459 unaffected parents, and 9 other unaffected family members; this cohort included 564 females and 537 males. The exome sequence data were from a subset of this cohort: 158 families consisting of 182 affected individuals, 150 unaffected siblings, 313 unaffected parents, 326 females, and 319 males. NIH, National Institutes of Health; PharmGKB, Pharmacogenomics Knowledgebase; SNP, single-nucleotide polymorphism.

recalibrated and genotyped by HaploTypeCaller according to GATK Best Practices using GATK v2.5-2 (refs. 8 and 9).

Variant annotation

Variants were annotated using transcript data from UCSC Genome Browser Database with ANNOVAR.¹⁰ The VCF files of SNP and exome sequencing variants in the study cohort were also annotated using the ClinVar database. The annotations taken from ClinVar included the dbSNP reference numbers, PubMed links, and PharmGKB annotation. The relationship data set from PharmGKB was used to annotate each variant with the associated medication names (Figure 1).

Pharmacological analysis

Using the annotated data, we identified patient SNPs with pharmacological implications or pharmacogenomic incidental findings. The number of variants with pharmacological implications that matched a patient's genotype was calculated per patient for both SNP genotypes and exome variants. The number of medications with potential implications was also calculated per patient.

Medication lists were compiled for study participants by extracting medication records from the NIH Biomedical Translational Research Information System and the NIH UDP Integrated Collaboration System. This medication list was then intersected with the list of medications associated with each participant's sequence variants. For each intersection, the validity of the medication match and subject's genotype, race, and sex were manually curated against the PharmGKB database and supporting publications.

Phenotype analysis

If available, patient medical records were reviewed to determine whether a patient had a medical history or phenotypic characteristics that correlated with the pharmacogenetic finding.

RESULTS

The PharmGKB database includes 868 annotated SNPs with published pharmacological implications. PharmGKB has six levels of evidence based on published evidence and the drug-dosing guidelines of the CPIC.¹¹ The top two levels of PharmGKB variants (1A or 1B category) have substantive evidence for clinical relevance. Level 1A variants have annotation for a variant-drug combination in a CPIC or medical society-endorsed PGx guideline. Level 1B variants have annotations for a variant-drug combination because the preponderance of evidence shows an association. The association for a Level 1B variant must be replicated in more than one cohort with significant p-values, and preferably will have a strong effect size.

The SNP chips used in our study provided 46% coverage of the SNPs annotated in the PharmGKB database and 53% coverage of high-priority SNPs (PharmGKB 1A and 1B categories). Variants identified through whole-exome sequencing conducted on the patients in the UDP covered 45% of the SNPs in the PharmGKB database and 58% of PharmGKB 1A and 1B SNPs. Combining the SNPs from both the SNP chip genotyping and

exome sequencing covered 65% of the SNPs in the PharmGKB database and 81% of the PharmGKB 1A and 1B SNPs.

All UDP patients with SNP genotypes or exome sequencing data had potential pharmacogenetic incidental findings. Within the SNP chip data, there were 696 potential incidental findings per patient that were associated with 276 different drugs. These included 19 variants ranked within PharmGKB categories 1A and 1B, and these 19 variants were associated with the efficacy of 17 drugs. Within the exome sequencing data, there were 728 detectable variants per patient that were associated with 283 different drugs related to 388 PharmGKB-documented SNPs. These included 21 variants within PharmGKB categories 1A and 1B that were associated with 14 different drugs. Combined SNP chip and exome data detected 949 variants per patient, and 29 were variants ranked in the PharmGKB 1A and 1B categories.

To determine whether these variants constituted incidental findings, i.e., were of therapeutic relevance, we focused on the 359 individuals for whom we had medication records. None of these individuals had been prescribed a medication for which they carried a PharmGKB 1A or 1B category variant. We therefore tested whether any of the pharmacogenetic variants they carried might provide insight into their response to a prescribed medication. This identified five pharmacogenetically relevant variants or incidental findings among nine individuals (Table 1). Three of the variants were detectable by both the SNP chip and exome sequencing, and two were detectable only by exome sequencing.

The five variants were relevant to efficacy of four different medications (Table 1): escitalopram, carbamazepine, morphine, and acetylsalicylic acid. First, a woman (patient 1) prescribed escitalopram had a PharmGKB category 3 association, *rs6318* genotype CG, which has been associated with altered function of the X-linked serotonin 5-HT_{2C} receptor and various phenotypes ranging from altered cortisol response to stress¹² to increased risk for cardiovascular disease mortality and morbidity.¹³ Second, a child (patient 2) who had seizures refractory to carbamazepine had a PharmGKB category 2B association, *rs1051740* genotype CT, which has been associated with a requirement for increased dosage of carbamazepine.¹⁴ Third, a woman (patient 3) who had been prescribed morphine and reported that ibuprofen gave better pain control had a PharmGKB category 3 association, *rs1799971* genotype AG, which has been associated with decreased efficacy of morphine in Caucasians.¹⁵⁻¹⁷ Fourth, a man (patient 8) who had cardiomyopathy and was prescribed acetylsalicylic acid had a PharmGKB category 3 association, *rs1799983* genotype GT, which has been associated with acetylsalicylic acid promoting in-stent restenosis.¹⁸ Fifth, a man (patient 9) with a history of myocardial infarction was prescribed acetylsalicylic acid and had a PharmGKB category 3 association, *rs5985* genotype AC, which has been associated with increased effectiveness of acetylsalicylic acid inhibition of factor XIII activation, ultimately lowering an individual's risk of myocardial infarction and death.¹⁹

DISCUSSION

Using the SNP chip data to analyze genotypes for 696 polymorphic variants and exome sequence data to define genotypes for

Table 1 Medically relevant pharmacogenomics variants in the NIH UDP cohort

UDP patient	SNP	Genotype	Age, sex, race	Phenotype	Drug	PharmGKB rank	Pharmacological implication
1	rs6318	CG	46, Female, Caucasian	Fibromyalgia, depression, hypertension	Escitalopram	3	CG associated with greater efficacy for treatment of nerve pain in males
2	rs1051740	CT	14, Male, Caucasian	Episodic ataxia type 1, seizures	Carbamazepine	2B	CT associated with increased metabolism of carbamazepine in people with epilepsy
3	rs1799971	AG	27, Female, Caucasian	Facial and abdominal pain, lung nodules, facial masses, recurrent pericarditis and pleuritis	Morphine	3	AG associated with increased metabolism and decreased efficacy for pain reduction
4	rs5985	AC	62, Female, Caucasian	Autonomic dysfunction, ataxia	Acetylsalicylic acid	3	Associated with inhibition of FXII activation
5	rs1799983	GT	57, Male, Caucasian	Neurological, no cardiovascular disease risks	Acetylsalicylic acid	3	Associated with increased in-stent restenosis in men with coronary artery disease
6	rs1799983, rs5985	GT, AC	52, Male, Caucasian	Fevers, CSF pleocytosis	Acetylsalicylic acid	3	Associated with increased in-stent restenosis in men with coronary artery disease and inhibition of FXII activation
7	rs5985	AC	68, Male, Caucasian	Fever, pulmonary	Acetylsalicylic acid	3	Associated with inhibition of FXII activation
8	rs1799983	GT	53, Male, Caucasian	Cardiac, heart murmur, aortic valve replacement, myopathy	Acetylsalicylic acid	3	Associated with increased in-stent restenosis in men with coronary artery disease
9	rs5985	AC	60, Male, Caucasian	Myopathy, fatigue, myocardial infarction, bicuspid aortic valve, pulmonary embolism	Acetylsalicylic acid	3	Associated with inhibition of FXII activation

CSF, cerebrospinal fluid; FXII, factor XIII; NIH, National Institutes of Health; PharmGKB, Pharmacogenomics Knowledgebase; SNP, single-nucleotide polymorphism; UDP, Undiagnosed Diseases Program.

728 polymorphic variants, we were able to interrogate the genotype of 949 pharmacogenetically relevant variants for each individual enrolled in the NIH UDP. Focusing on the 359 individuals for whom we had medication information, we discovered five reportable variants among nine individuals from nine different families (eight probands and one unaffected family member). For five (56%) of these individuals, the pharmacogenetic incidental findings were potentially relevant to their medical management.

Although the sequencing conducted by the NIH UDP was not designed to screen pharmacologically relevant variants, 65% of the annotated variants in the PharmGKB databases were represented in the UDP sequencing data. All participants in the study received SNP chip genotyping; exome sequencing was conducted for 59% of these individuals. Both SNP chip genotyping and exome sequencing were important for identifying the pharmacogenetically relevant variants. Although 50% of the variants that were discovered in the UDP dataset could be identified with either the SNP chip genotyping data or the exome sequencing data, 23% were found only within the SNP chip genotyping data and 27% were found only within the exome sequencing data.

Four additional issues arising during our analysis were (i) defining the level of relative risk warranting reporting of a potential variant, (ii) determining how to weight variants identified in an ethnic group different from that of the subject, (iii) the need for clinical correlation, and (iv) obligations to family members. Relevant to the first issue, clinicians expressed a need for prioritization of the findings because there were a large number of potential incidental findings per patient. To address this request, we used the clinical annotation levels of evidence of PharmGKB. Additionally, we reported the ethnicity within which the variant–drug associations were identified and highlighted any differences in ethnicity between the patient and that population and prioritized variant–drug associations accordingly.

Although filtering the incidental findings for PharmGKB 1A and 1B categories produces a manageable list of 21 medications, none of the individuals for whom we had medication data had a PharmGKB category 1A or 1B association. We therefore tested for associations ranked lower in PharmGKB and identified five PharmGKB category 2 or category 3 associations. Five of these were potentially relevant to the care of the individuals, three patients, and one family member (Table 1).

A one-sentence summary of the PharmGKB variant with the associated drug information was incorporated into the report entered into the NIH UDP database. Entry of the findings into the database also enabled sorting of the results and searching by individual, drug name, and priority. Primary clinicians’ discussions regarding whether to inform individuals enrolled under the NIH UDP protocol about the identified variants focused on the delineated and perceived obligations defined by the language of the consent document and the process by which the consent was explained. The choice of whether to report a given variant to a given study subject was deferred to the relevant clinical team. No clinical team elected to return a PharmGKB

category 2 or 3 variant detected by this study for at least two reasons. First, the study consent reflected routine practice from the early days of the application of genome-scale sequencing to medical diagnostics—only DNA variants that might contribute to the test indication were to be returned. Second, current ACMG guidelines do not include such variants among those recommended to be returned.² The UDP recognizes that these are areas of intense debate in the literature and elsewhere; the program is prepared to adjust its practice as the standard of care evolves.

Our analysis had some limitations. Some of the variant annotations in PharmGKB associate a variant to a class of drugs (e.g., bisphosphonates or antidepressants) instead of a specific drug. In contrast, UDP medication records list the specific drug name; therefore, our simple matching algorithm did not identify all associations. This problem can be readily addressed with drug ontology that would allow drug class matches. We also did not have access to the complete list of medications administered to our patients prior to their participation in our study. Despite these limitations, a small percentage of individuals with relevant medication prescriptions were identified in the cohort, indicating that there was probably an even larger percentage of medically relevant findings if the computational filtering mechanisms were refined.

The results of this study indicate that, by looking for pharmacogenetically relevant findings in SNP genotypes or exome sequencing data, incidental findings can be identified and potentially provide a valuable resource for patient care. The interpretation, use in medical practice, and returning of pharmacogenomics secondary findings are, however, different than those for the nonpharmacogenomic incidental findings listed in the original ACMG publication.² These pharmacogenomic variants are different from nonpharmacogenomic incidental findings in that they are relevant only in the presence of an environmental factor, i.e., the drug that they have an effect on. This means that a pharmacogenomic variant may never be relevant in one person's lifetime but may be absolutely important for another person who is prescribed the drug in question. For example, in a recent study of 48 pharmacogenomically relevant variants among a cohort of 94 patients,²⁰ the electronic medical record decision support system generated no alerts during the hospital stay of these patients.

Nonetheless, our described methodology and findings suggest that identifying and reporting pharmacogenetic incidental findings can improve patient care and personalized medicine. Additionally, if medically relevant findings can be found in the small UDP cohort with sequencing strategies designed for diagnostic purposes and not pharmacogenetic purposes, then we hypothesize that such incidental findings occur in the sequencing data for other medical programs.

In summary, using SNP chip genotyping and exome sequencing, the NIH UDP identified pharmacogenetic incidental findings in its cohort. We have established a framework for evaluating and selecting pharmacogenomic variants that are potentially useful for therapeutic management.

ACKNOWLEDGMENTS

We thank NHGRI Intramural Sequencing Center and Appistry, Inc., for their sequencing, alignment, genotyping, and annotation services. This work was supported in part by the Common Fund, Office of the Director, and the Intramural Research Program of the National Human Genome Research Institute (NIH, Bethesda, MD, USA).

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. Wolf SM, Crock BN, Van Ness B, et al. Managing incidental findings and research results in genomic research involving biobanks and archived data sets. *Genet Med* 2012;14:361–384.
2. Green RC, Berg JS, Grody WW, et al.; American College of Medical Genetics and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 2013;15:565–574.
3. Leeder JS. Developmental and pediatric pharmacogenomics. *Pharmacogenomics* 2003;4:331–341.
4. Hawcutt DB, Thompson B, Smyth RL, Pirmohamed M. Paediatric pharmacogenomics: an overview. *Arch Dis Child* 2013;98:232–237.
5. Gahl WA, Tifft CJ. The NIH Undiagnosed Diseases Program: lessons learned. *JAMA* 2011;305:1904–1905.
6. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 2009;84:210–223.
7. Rozowsky J, Abyzov A, Wang J, et al. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol* 2011;7:522.
8. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–498.
9. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–1303.
10. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
11. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther* 2011;89:464–467.
12. Brummett BH, Babyak MA, Williams RB, et al. A putatively functional polymorphism in the HTR2C gene is associated with depressive symptoms in white females reporting significant life stress. *PLoS One* 2014;9:e114451.
13. Brummett BH, Babyak MA, Jiang R, et al. A functional polymorphism in the 5HTR2C gene associated with stress responses also predicts incident cardiovascular events. *PLoS One* 2013;8:e82781.
14. Makmor-Bakry M, Sills GJ, Hitiis N, Butler E, Wilson EA, Brodie MJ. Genetic variants in microsomal epoxide hydrolase influence carbamazepine dosing. *Clin Neuropharmacol* 2009;32:205–212.
15. Campa D, Gioia A, Tomei A, Poli P, Barale R. Association of ABCB1/MDR1 and OPRM1 gene polymorphisms with morphine pain relief. *Clin Pharmacol Ther* 2008;83:559–566.
16. Matic M, Simons SH, van Lingen RA, et al. Rescue morphine in mechanically ventilated newborns associated with combined OPRM1 and COMT genotype. *Pharmacogenomics* 2014;15:1287–1295.
17. Reyes-Gibby CC, Shete S, Ravvåg T, et al. Exploring joint effects of genes and the clinical efficacy of morphine for cancer pain: OPRM1 and COMT gene. *Pain* 2007;130:25–30.
18. Shuvalova YA, Kaminnyi AI, Meshkov AN, Shirokov RO, Samko AN. Association between polymorphisms of eNOS and GPx-1 genes, activity of free-radical processes and in-stent restenosis. *Mol Cell Biochem* 2012;370:241–249.
19. Undas A, Sydor WJ, Brummel K, Musial J, Mann KG, Szczeklik A. Aspirin alters the cardioprotective effects of the factor XIII Val34Leu polymorphism. *Circulation* 2003;107:17–20.
20. Nishimura AA, Shirts BH, Dorschner MO, et al. Development of clinical decision support alerts for pharmacogenomic incidental findings from exome sequencing. *Genet Med* 2015;17:939–942.