Published in final edited form as:

Pharmacogenet Genomics. 2016 April ; 26(4): 161–168. doi:10.1097/FPC.0000000000000202.

PGRNseq: A Targeted Capture Sequencing Panel for Pharmacogenetic Research and Implementation

Adam Gordon¹, Robert S. Fulton³, Xiang Qin², Elaine R. Mardis³, Deborah A. Nickerson^{1,*}, and Steve Scherer²

¹Department of Genome Sciences, University of Washington

² The Baylor College of Medicine Human Genome Sequencing Center, Houston, TX

³The Genome Institute at Washington University, St. Louis, MO

Abstract

Objectives—While the costs associated with whole-genome and whole-exome next-generation sequencing continue to decline, they remain prohibitively expensive for large-scale studies of genetic variation. As an alternative, custom-target sequencing has become a common methodology based on its favorable balance between cost, throughput, and deep coverage.

Methods—We have developed PGRNseq, a custom-capture panel of 84 genes with associations to pharmacogenetic phenotypes, as a tool to explore the relationship between drug response and genetic variation, both common and rare. We utilized a set of 32 diverse HapMap trios and 2 clinical cohorts to assess platform performance, accuracy, and ability to discover novel variation.

Results—We find that PGRNseq generates ultra-deep coverage data (mean = 496x) that is over 99.8% concordant with orthogonal datasets. Additionally, in our testing sets, PGRNseq identified many novel, rare variants of interest, underscoring its value in both research and clinical settings.

Conclusion—PGRNseq is an ideal platform for performing sequencing-based analyses of pharmacogenetic variation in large cohorts. Additionally, the high accuracy associated with genotypes from PGRNseq highlight its utility as a clinical test.

Keywords

Pharmacogenomics; Targeted Sequencing; Rare Variation

Introduction

As next-generation sequencing costs continue to decrease, and rare variant analysis becomes an imperative, sequencing-based association analysis is developing as a widely applied tool in human genetic analysis through whole exome and whole genome sequencing as well as the application of targeted sequencing panels. Indeed, these approaches have been successful in identifying novel associations between genetic variation and a range of traits including cardiovascular, psychiatric, and pharmacogenetic phenotypes [1-3]. However, sequencing

*Corresponding author: debnick@uw.edu.

full genomes or even full exomes of the tens of thousands of individuals needed for adequately powered association studies remains costly and time-consuming. Targeted high-throughput sequencing panels, which capture and sequence a small set of genomic targets to high depth, represent a middle-ground that maximizes throughput while maintaining the deep coverage characteristic of high-quality next generation sequencing (NGS) data. To date, targeted sequencing panels have been successfully deployed in both clinical research and diagnostics with applications as diverse as the mutational analysis of individuals with Lynch or polyposis syndrome [4], the detection of somatic mutation in lung cancer [5], and the molecular diagnosis of retinitis pigmentosa [6]. In this article we will discuss the process of creating and validating such a panel, focusing on 84 genes of pharmacogenetic importance, including many genes identified as actionable by the Clinical Pharmacogenetics Implementation Consortium (CPIC) [7].

Since its inception, pharmacogenetic research has identified many genes that play a role in drug metabolism, transport, and response, and has shown that many variants within these genes contribute to overall variation in drug phenotypes [8]. These gene-phenotype pairs include drug-metabolizing enzymes (such as *CYP2C19* and clopidogrel response [9]), drug transporters (such as *UGT1A1* and irinotecan [10]) and specific drug targets (such as *VKORC1* and warfarin [11]). Many of these drug-gene pairs are clinically actionable, and the number of clinical entities performing pharmacogenetic testing is increasing steadily. Despite this increasing popularity, most of the known variation within these pharmacogenes is common (i.e. MAF > 1%). Indeed, many existing platforms are currently used to genotype these targets in a high-throughput manner, such as the Affymetrix DMET+ array and the Illumina ADME assay which focus largely on common variation. However, initial large-scale, NGS-based studies have revealed that rare (i.e. MAF < 1%) deleterious variation is in fact collectively common across drug metabolizing enzyme and drug targets; though each individual variant can be vanishingly rare. In fact, 7-10% of individuals harbor such a variant [12]. Additionally, rare variation in pharmacogenes has been directly linked to variation in drug response and to rare adverse events in the several cases that have been studied extensively to date [13,14]. Thus, it's clear that this category of variation is of importance as pharmacogenetic testing expands clinically. This necessitates collaborative efforts on the analysis of rare variation in pharmacogenes.

The Pharmacogenomics Research Network (PGRN) is a collaborative network formed in order to coordinate pharmacogenetic research and to collectively provide recommendations as to the clinical relevance of pharmacogenetic variation. Seeing an opportunity to facilitate large-scale sequencing studies of pharmacogenetic targets to assess both rare and common variation, as well as an opportunity to explore the clinical utility of NGS, the PGRN called on the network's 3 Deep Sequencing Resources (DSRs: Department of Genome Sciences, University of Washington (UW); The Genome Institute at Washington University (WashU); and the Human Genome Sequencing Center at Baylor College of Medicine (BCM-HGSC)) to develop a custom-capture panel centered on pharmacogenes of known interest. Here we present an overview of the design, testing, and quality control of PGRNseq; the resulting panel is currently available for use by members of the pharmacogenetics community.

Methods

PGRNseq Design

PGRN network members nominated genes for consideration in the design of the NGS platform. As one of the design criteria was to produce a panel that could be cost competitive with genotyping arrays, not all nominations could be included in the final list. Through multiple rounds of balloting and discussion, the group collectively decided on a final consensus list of 84 pharmacogenes for inclusion in the panel (Table 1). These pharmacogenes are functionally diverse and include drug-metabolizing enzymes, drug transporters, and drug targets. Although all 84 genes have some prior association with a pharmacogenetic trait, they range from those deemed clinically actionable by CPIC [7] at the time of voting to those about which little is known aside from strong preliminary association data. For the design of each of the 84 genes, we included all exons (based on all transcript models) as well as 2kb upstream and 1kb downstream of their untranslated regions (UTRs) in order to discover and assess nearby potential regulatory variation, which is already known to affect drug response in genes such as *VKORC1* [15]. In addition, the design also included probes to capture every site present on the Affy DMET+ array and the Illumina ADME assay, so that the sequencing platform would be backwards compatible with existing datasets, and as orthogonal platforms for PGRNseq quality control via genotype concordance. After submitting the final list of genomic coordinates to Nimblegen, we worked closely with their developmental team to generate a set of probes to capture these regions; the resulting set of SeqCap probes, known as PGRNseq, covers 968kb of the genome, which is highly scalable for large studies while maintaining high coverage.

PGRNseq Testing

In order to test different multiplexing strategies and assess the accuracy of the platform, the DSRs assembled a set of 96 HapMap samples of diverse ancestry (HapMap96). Since all samples have HapMap genotypes available, and some have 1000 Genomes sequencing data available, they represent a robust set to assess overall platform performance and concordance. Furthermore, these 96 samples consist of 32 trios, so analysis of Mendelian inheritance can reveal sites prone to false-positive calls due to mapping errors deriving from repetitive elements or from regions of high sequence homology; several genes on the platform are members of large gene families that can be prone to these errors. In addition to these samples, we also wanted to test PGRNseq performance on cohorts of actual patients with orthogonal data. We obtained two different clinical cohorts: 1) a set of 96 liver-derived patient samples [16], and 2) a separate set of 96 clinical samples collected for research into antiplatelet response for testing.

All sequencing was performed on the Illumina HiSeq 2000 instrument using paired-end, 100bp reads. Initially, all three DSR groups tested the HapMap96 using a variety of capture probe and sequencing lane multiplexing strategies (8-plex, 12-plex, 24-plex) to identify the maximum batch size that preserves the sequence read depth needed for high quality variant calls across the target set; with these criteria in mind the group settled on a 24-plex batch size. To compare performance, Illumina ADME genotypes and Affy DMET+ genotypes were generated for the HapMap96 at UW and BCM-HGSC, respectively. Clinical cohorts

were sequenced at UW (liver samples) and BCM-HGSC (antiplatelet samples) using the same protocol as was used for the HapMap96 assays. At each site, raw sequencing reads were mapped to the hg19 reference genome using the Burrows-Wheeler Aligner (BWA), and variants called and filtered using GATK [17] and ATLAS [18].

Results

General Platform Performance

Using a 24-plex capture strategy leads to an average coverage of 496X across the target space, demonstrating that PGRNseq can consistently generate ultra-deep sequencing data while maintaining the high throughput necessary for studies of large sample size (Table 2). At the single gene level, PGRNseq generates deep coverage data for the complete coding region for nearly every captured gene (Table 1). The major exceptions are the two MHC genes on the platform, *HLA-B* and *HLA-DQB3* despite the inclusion of all 8 alternative reference haplotypes in the design phase. As these genes are highly structurally polymorphic, they present a considerable challenge to assemble using short reads [19]. Therefore, SNV calls in this region were not considered further. Other areas of low coverage (mean depth < 20X) consist largely of noncoding regions distant from the coding regions, and in most cases these low coverage regions were also related to the presence of repetitive elements within the 2kb/1kb upstream/downstream design window.

Quality and accuracy of PGRNseq variants

Although PGRNseq generates high-coverage data many targeted genes could be prone to erroneous variant calls due to sequence homology with other gene family members and the associated inappropriate capture and/or sequence read mismapping, or due to the presence of structural variants (SVs). Therefore, we assessed accuracy of PGRNseq genotype calling using orthogonal datasets as well as consistency with Mendelian inheritance. Across the HapMap96, we observed an average of 1325 total variants called per individual. See Supplemental Content 1 for counts of variants per gene separated by variant type. Analysis of Mendelian inconsistencies within the 32 trios revealed that the majority of genes (63/82, 77%) did not contain any such errors (Supp. Content 1). Of the 19 genes that did, 17 contained 3 or fewer Mendelian errors, all of which were in noncoding regions at the edges of the target space. However, two genes, *CYP2A6* and *CYP2D6*, contained 10 or more Mendelian inconsistencies. These results were not unexpected as both genes have one or more neighboring pseudogenes with high homology, and both are known to harbor structural variants of functional consequence [20,21]. Indeed, many of these inconsistencies were found across multiple trios and located in regions of low unique mappability (Supplemental Content 2).

In addition to the quality checks inherent in the use of trio data, we chose a panel of HapMap samples in order to compare our results to those from other large sequencing or genotyping efforts, e.g. 1000 Genomes. To evaluate accuracy, we calculated the mean per-individual genotype concordance at various coverage cutoffs using 3 different datasets: HapMap 3.3 (n=96) [22], 1000 Genomes deeply sequenced trios (n=6) [23], and high-coverage exome data generated at UW through the NIEHS Environmental Genome Project

(n=54) [24]. Generally, mean per-individual concordance was greater than 99% (Table 3). We noticed that the mean of 99.4% concordance with HapMap 3.3 was consistent across different depth cutoffs; and on closer analysis, we found that 2 noncoding sites were solely responsible for these discrepancies. We believe these sites may be difficult to type using chip-based genotyping, as genotypes from the three different sequencing datasets agree with the PGRNseq genotype at this site. In fact, the two sequencing-based comparison datasets (1000 Genomes deep trios and EGP exomes) showed mean per-individual concordance greater than 99.8%, indicating that the vast majority of PGRNseq-derived genotypes are accurate.

We observed a similarly high accuracy when PGRNseq genotypes were compared to those derived from the Pharmacogenetics-specific genotyping platforms, Affy DMET+ and Illumina ADME. Specifically, the average per-individual concordance was 99.7% between PGRNseq genotypes and Affy DMET+ data (Supp. Content 3). When comparing the Illumina ADME data, 87 of the HapMap96 were 100% concordant across all overlapping sites. The remaining 9 individuals were all discordant for a single shared variant. As PGRNseq genotypes at this site (rs9282861) are concordant with genotypes for these same individuals derived from HapMap 3.3 and from the other sequencing-based comparison sets, this discordance likely arose either from an underlying CNV in these individuals or inaccurate cluster calling by the Illumina ADME platform (Supp. Content 4). Based on these data, sequencing-based PGRNseq genotypes are exceedingly accurate.

Novel variation in the HapMap96 and clinical cohorts

Across 82 genes on the panel (excluding the MHC genes), we identified an average of 45 variants per HapMap96 individual that were not present in dbSNP build 137. This value is similar to data from the liver cohort (mean = 35 novel variants per individual) and the antiplatelet cohort (mean = 55 novel variants per individual), which consisted largely of Caucasian individuals, and were less diverse than the trios in the HapMap96. Though the majority of novel variants were in noncoding regions, we identified several novel, potentially deleterious nonsense and missense variants across both the HapMap96 and clinical cohorts within genes of clinical importance such as *RYR1*, *CYP2C9*, and *SLCO1B1*, (Supp. Content 1, 5) [25,26]. As the sample sizes of the testing cohorts are relatively small, we conclude that PGRNseq can identify many more novel alleles of interest when applied to large studies that are well-powered to detect rare variation associated with variation in drug response.

Discussion

As adverse drug reaction events are a significant cause of morbidity in the US [27], a platform that can accurately detect and genotype variants, both common and rare, that affect drug response has the potential to both deepen our understanding of these events as well as reduce their incidence. As recent large scale sequencing efforts have revealed that very rare, potentially deleterious variants are carried by nearly 1 in 10 individuals [12]. Therefore, platforms like PGRNseq that can accurately detect such variation, as well as genotype common variants of known effect, are particularly well-suited for pharmacogenetic research and clinical implementation. To this end, we have developed a novel custom-target panel

designed to capture 84 genes with known roles in drug metabolism and response phenotypes. Aimed at larger pharmacogenomics studies, PGRNseq strikes a favorable balance between low cost, the high throughput associated with chip-based genotyping and the ultra-deep coverage associated with NGS; additionally, the deep coverage inherent to quality NGS data enables the discovery of rare variation of potential clinical impact. The cost of sequencing a sample using the PGRNseq platform is less expensive than other known methods to survey the comparable space. Although costs vary from institution to institution, generally the PGRNseq platform is about eight to ten times cheaper than whole genome sequencing, and two to three times cheaper than exome sequencing. In addition to the data generation savings, the data storage and data analysis costs are also much less and more efficient. Genotyping arrays, although comparably priced, do not assay the entire gene space that the PGRNseq product does, and thus are inferior to the sequence data produced by PGRNseq. Testing PGRNseq across multiple sample sets revealed the high accuracy of genotypes obtained from this platform. Finally, the quantity and types of novel variation discovered in our small testing sets demonstrates that PGRNseq is ideal for larger sequencing studies aimed at assessing rare variation within pharmacogenomics targets. Indeed, several current studies are making use of the PGRNseq platform. These studies range from investigations into a specific drug response phenotypes, such as irinotecan response in cancer patients, to large, multi-institution sequencing efforts such as eMERGE-PGx, which is deploying PGRNseq prospectively across 9000 patients with linked medical records in order to discover rare pharmacogenetic alleles of interest and to explore the utility of clinical NGS implementation [28]. Overall, we believe PGRNseq will continue to be a valuable resource for any investigator interested in examining rare variation in targets of known pharmacogenomic importance.

Although PGRNseq is currently being deployed in studies such as those described above, we are also in the process of designing and testing a ‘version 2.0’ platform to expand our abilities to target variation within the complex regions on the platform: *CYP2D6*, *CYP2A6*, and *HLA-B*. For *CYP2D6* and *CYP2A6*, we intend to extend the probe design to include the full gene (introns included) as well as the neighboring pseudogenes in order to aid in the assembly of this complex region, particularly with longer read lengths. Although generally well-captured, the assembly and interpretation in a clinical or research setting of data from these regions, subject to complex structural rearrangements [29], requires the development of new computational resources. We have also identified for inclusion several variants that are well known to tag the two “actionable” *HLA-B* alleles (*57:01 and *15:02) and that can be accurately typed by NGS [30]. Finally, we also plan to fine-tune our coverage of non-coding space by focusing on regions of putative regulatory function and removing low-coverage, repetitive regions that happened to fall within the boundaries of the original design scheme.

In addition to PGRNseq development, we feel the design and testing strategy utilized here is broadly applicable to the development of any custom-capture panel focusing on specific subsets of genomic targets. The use of the HapMap96 was essential in assessing optimum platform conditions, overall performance, and genotype accuracy due to the abundance of orthogonal data on these samples as well as the Mendelian inheritance analysis enabled by the use of trios, specifically. Though initially intended as a research tool, our experiences in

the design and testing of this specific platform has led to an interest in pursuing the use of PGRNseq as a clinical test, and efforts to clinically validate this platform for certain actionable alleles are currently underway. As custom target platforms such as PGRNseq continue to demonstrate their efficacy as a research tool for the study of genetic variation, both rare and common, we believe that these very same platforms will become the standard for clinical sequencing, carving a translational niche for NGS in medicine ahead of clinical whole-genome sequence implementation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors wish to acknowledge the Pharmacogenomics Research Network (PGRN) for supporting the development of the PGRNseq platform as well as the contributions of Richard Wilson from the Genome Institute at Washington University and Richard Gibbs and Donna Muzny from the Human Genome Sequencing Center at Baylor College of Medicine. This work was supported by the following grants: GM061388, HL069757 and GM097119.

References

1. The TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *New Engl J Med*. 2014; doi: 10.1056/NEJMoa1307095
2. O'Roak BJ, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*. 2012; 485:246.doi: 10.1038/nature10989 [PubMed: 22495309]
3. Tammiste A, Jiang T, Fischer K. Whole-Exome Sequencing Identifies a polymorphism in the BMP5 Response Gene associated with SSRI treatment in Major Depression. *J Psychopharm*. 2013; (10): 915–920.
4. Pritchard CC, Smith C, Salipante SJ, Lee MK, Thornton AM, et al. ColoSeq provides comprehensive Lynch and polyposis syndrome mutational analysis using massively parallel sequencing. *J Mol Diagn*. 2012; 14:357–366. DOI: 10.1016/j.jmoldx.2012.03.002 [PubMed: 22658618]
5. Kim EH, Lee S, Park J, Lee K, Bhak J, Kim BC. New lung cancer panel for high- throughput targeted resequencing. *Genomics Inform*. Jun; 2014 12(2):50–7. DOI: 10.5808/GI.2014.12.2.50 [PubMed: 25031567]
6. Wang F, Wang H, Tuan HF, Nguyen DH, et al. Next generation sequencing-based molecular diagnosis of retinitis pigmentosa: identification of a novel genotype phenotype correlation and clinical refinements. *Hum Genet*. Mar; 2014 133(3):331–45. DOI: 10.1007/s00439-013-1381-5 [PubMed: 24154662]
7. Caudle KE, Klein TE, Hoffman JM, Muller DJ, Whirl-Carrillo M, et al. Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline development process. *Curr Drug Metab*. Feb; 2014 15(2):209–17. [PubMed: 24479687]
8. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science*. 1999; 286:487–491. [PubMed: 10521338]
9. Scott SA, Sangkuhl K, Gardner EE, Stein CM, Hulot JS, Johnson JA, Roden DM, Klein TE, Shuldiner AR. Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450-2C19 (CYP2C19) genotype and clopidogrel therapy. *Clin. Pharmacol. Therapeut*. 2011; 90(2): 328–32.

10. Liu X, Cheng D, Kuang Q, Liu G, Xu W. Association of UGT1A1*28 polymorphisms with irinotecan-induced toxicities in colorectal cancer: a meta- analysis in Caucasians. *Pharmacogenomics J.* 2013
11. Johnson JA, Gong L, Whirl-Carrillo M, Gage BF, Scott SA, Stein CM, Anderson JL, Kimmel SE, Lee MT, Pirmohamed M, Wadelius M, Klein TE, Altman RB. Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. *Clin. Pharmacol. Therapeut.* 2011; 90(4):625–9.
12. Gordon AS, Tabor HK, Johnson AD, Snively BM, Assimes TL, Auer PL, Ioannidis JP, Peters U, Robinson JG, Sucheston LE, et al. Quantifying rare, deleterious variation in 12 human cytochrome p450 drug- metabolism genes in a large-scale exome dataset. *Hum. Mol. Gen.* 2014; 23:1957–1963. DOI: 10.1093/hmg/ddt588 [PubMed: 24282029]
13. Ramsey LB, Bruun GH, Yang W, Trevino LR, Vattathil S, Scheet P, et al. Rare versus common variants in pharmacogenetics: SLCO1B1 variation and methotrexate disposition. *Genome Res.* 2012; 22:1–8. DOI: 10.1101/gr.129668.111 [PubMed: 22147369]
14. Ramirez AH, Shaffer CM, Delaney JT, Sexton DP, Levy SE, Rieder MJ, Nickerson DA, George AL Jr. Roden DM. Novel rare variants in congenital cardiac arrhythmia genes are frequent in drug-induced torsades de pointes. *Pharmacogenomics J.* 2013; 13:325–329. DOI: 10.1038/tpj.2012.14 [PubMed: 22584458]
15. Wang D, Chen H, Momary KM, Cavallari LH, Johnson JA, Sadée W. Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (VKORC1) affects gene expression and warfarin dose requirement. *Blood.* Aug 15; 2008 112(4):1013–21. [PubMed: 18523153]
16. Innocenti F, Cooper GM, Stanaway IB, Gamazon ER, et al. Genetic identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLOS Genetics.* 2011; 7(5):e1002078. [PubMed: 21637794]
17. Van der Auwera GA, Carneiro M, Hartl C, Poplin R, del Angel G, Levy- Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella K, Altshuler D, Gabriel S, DePristo M. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Pro. Bioinf.* 2013; 43
18. Challis D, Yu J, Evani US, Jackson AR, et al. An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics.* Jan 12.2012 13:8. [PubMed: 22239737]
19. Major E, Rigó K, Hague T, Bérces A. Juhos S: HLA Typing from 1000 Genomes Whole Genome and Whole Exome Illumina Data. *PLoS One.* 2013; 8(11):e78410. [PubMed: 24223151]
20. Hicks JK, Swen JJ, Gaedigk A. Challenges in CYP2D6 phenotype assignment from genotype data: a critical assessment and call for standardization. *Curr Drug Metab.* Feb; 2014 15(2):218–32. [PubMed: 24524666]
21. Mwenifumbo JC, Tyndale RF. Genetic variability in CYP2A6 and the pharmacokinetics of nicotine. *Pharmacogenomics.* 2007; 8(10):1385–1402. [PubMed: 17979512]
22. The International HapMap Consortium. The International HapMap Project. *Nature.* 2003; 426:789–796. [PubMed: 14685227]
23. 1000 Genomes Consortium. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012; 15(7422):56–65.
24. NIEHS Environmental Genome Project. Seattle, WA: URL: <http://evs.gs.washington.edu/niehsExome/>
25. Kim JH, Jarvik GP, Browning BL, Rajagopalan R, et al. Exome sequencing reveals novel rare variants in the ryanodine receptor and calcium channel genes in malignant hyperthermia families. *Anesthesiology.* 2013; 119(5):1054–1065. [PubMed: 24013571]
26. Wilke RA, Ramsey LB, Johnson SG, Maxwell WD, McLeod HL, Voora D, Krauss RM, Roden DM, Feng Q, Cooper-Dehoff RM, Gong L, Klein TE, Wadelius M, Niemi M. The clinical pharmacogenomics implementation consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. *Clin Pharmacol Ther.* Jul; 2012 92(1):112–7. [PubMed: 22617227]
27. Budnitz DS, Pollock DA, Weidenbach KN, Mendelsohn AB, Schroeder TJ, Annest JL. National surveillance of emergency department visits for outpatient adverse drug events. *JAMA.* 2006; 296(15):1858–66. [PubMed: 17047216]

28. Rasmussen-Torvik LJ, Stallings SC, Gordon AS. Design and Anticipated Outcomes of the eMERGE-PGx Project: A Multi-Center Pilot for Pre-Emptive Pharmacogenomics in Electronic Health Record Systems. *Clin Pharmacol Ther.* Jun 24.2014 doi: 10.1038/clpt.2014.137
29. Gaedigk A. Complexities of CYP2D6 gene analysis and interpretation. *Int Rev Psychiatry.* Oct; 2013 25(5):534–53. [PubMed: 24151800]
30. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet.* 2006; 38:1166–1172. DOI: 10.1038/ng1885 [PubMed: 16998491]

Table 1

84 Pharmacogenes of interest captured by PGRNseq and their overall performance.

Gene Symbol	Chromosome	Coding Plus length (bp)	Gene Function/Role	Mean HapMap96 coverage	Mean HapMap96 coverage (coding only)	% Coding bases >30×
ABCA1	9	9982	Target	350.24	397	100
ABCB1	7	7480	Absorption	286.73	415	100
ABCB11	2	7074	Absorption	335.36	458	100
ABCC2	10	7766	Absorption	355.52	567	100
ABCG1	21	5265	Absorption	392.68	372	98
ABCG2	4	5028	Absorption	250.18	522	100
ACE	17	7224	Target	378.92	275	96
ADRB1	10	4438	Target	252.14	539	79
ADRB2	5	4246	Target	356.08	329	100
AHR	7	5591	Metabolism	248.22	495	100
ALOX5	10	5081	Target	326.2	418	100
APOA1	11	3816	Target	370.24	420	100
ARID5B	10	6607	Disease	321.12	461	100
BDNF	11	3804	Target	311.84	565	100
CACNA1C	12	9936	Target	380.02	686	100
CACNA1S	1	8622	Target	439.26	385	100
CACNB2	10	5349	Target	271.8	490	100
CES1	16	4763	Metabolism	350.84	567	77
CES2	16	4920	Metabolism	346.27	605	100
COMT	22	3939	Metabolism	361	596	100
CRHR1	17	4300	Target	302.18	690	100
CYP1A2	15	4575	Metabolism	286.58	529	100
CYP2A6	19	6052	Metabolism	339.84	483	100
CYP2B6	19	4512	Metabolism	331.18	475	100
CYP2C19	10	5648	Metabolism	332.72	451	100
CYP2C9	10	4509	Metabolism	315.8	441	100
CYP2D6	22	4530	Metabolism	327.6	306	97

Gene Symbol	Chromosome	Coding Plus length (bp)	Gene Function/Role	Mean HapMap96 coverage	Mean HapMap96 coverage (coding only)	% Coding bases >30×
CYP2R1	11	4524	Metabolism	320.54	428	100
CYP3A4	7	4564	Metabolism	361.7	421	100
CYP3A5	7	4561	Metabolism	310.1	565	100
DBH	9	4854	Target	429.54	351	100
DPYD	1	6170	Excretion	292.67	505	100
DRD1	5	4345	Target	304.08	618	100
DRD2	11	4360	Target	390.34	500	100
EGFR	7	7009	Target	356.74	416	98
ESR1	6	5065	Target	318.08	432	99
FKBP5	6	4414	Target	315.66	415	100
G6PD	X	4690	Drug-induced Disease	269.4	285	96
GLCC11	7	4676	Drug-induced Disease	289.12	435	76
GRK4	4	4801	Target	340.84	579	100
GRK5	10	4837	Target	337.76	101	100
HLA-B	6	3000	Toxicity	98.64	132	13
HLA-DQB3	6	3000	Toxicity	123.6	368	22
HMGCR	5	5743	Target	295.04	365	100
HSD11B2	16	4238	Metabolism	358.96	351	78
HTR1A	5	4273	Target	284.2	430	100
HTR2A	13	4428	Drug-induced Disease	333.02	400	100
KCNH2	7	6921	Drug-induced Disease	306.64	597	87
LDLR	19	5655	Target	318.74	296	100
MAOA	X	4644	Target	239.4	333	100
NAT2	8	3877	Metabolism/Excretion	286	444	100
NPPB	1	3417	Drug-induced Disease	348.22	503	100
NPR1	1	6186	Target	336.04	378	98
NR3C1	5	5421	Target	301.53	408	100
NR3C2	4	5955	Target	294.06	454	100
NTRK2	9	5664	Target	330.66	515	100

Gene Symbol	Chromosome	Coding Plus length (bp)	Gene Function/Role	Mean HapMap96 coverage	Mean HapMap96 coverage (coding only)	% Coding bases >30×
PEAR1	1	6202	Target	337.84	514	100
POR	7	5103	Drug-induced Disease	343.6	576	100
PTGIS	20	4543	Target	381.64	604	97
PTGS1	9	4844	Target	393.76	522	100
RYR1	19	18541	Drug-induced Disease	392.65	383	97
RYR2	1	18324	Drug-induced Disease	320.3	579	100
SCN5A	3	9255	Drug-induced Disease	400.44	408	100
SLC15A2	3	5278	Excretion	297.8	552	100
SLC22A1	6	4709	Excretion	337.75	435	100
SLC22A2	6	4712	Excretion	351.12	332	100
SLC22A3	6	4715	Excretion	304.1	554	88
SLC22A6	11	4732	Absorption	332.01	538	100
SLC47A1	17	4781	Absorption	334.68	513	100
SLC47A2	17	4877	Absorption	402.04	665	100
SLC6A3	5	4919	Target	371.69	566	100
SLC6A4	17	4945	Disease	313.64	299	100
SLCO1A2	12	5476	Absorption	265.88	291	100
SLCO1B1	12	5132	Absorption	260.12	287	100
SLCO1B3	12	5165	Absorption	249.76	587	100
SLCO2B1	11	5186	Absorption	359.58	470	100
TBXAS1	7	4657	Metabolism	351.4	463	100
TCL1A	14	3357	Disease	379.86	336	100
TPMT	6	3770	Metabolism	251.13	290	100
UGT1A1	2	4622	Excretion	322.58	502	98
UGT1A4	2	6806	Excretion	352.58	611	100
VDR	12	4316	Absorption	391.52	580	100
VKORC1	16	3504	Target	308.46	626	100
ZNF423	16	6887	Target	405.36	626	100

These genes were nominated and voted on by the PGRN community for inclusion in the final target. "Coding Plus" length indicates the number of base pairs that make up the gene's exons as well as 2kb upstream and 1kb downstream of the coding region. Function/Role annotations derived from PharmGKB. Per-gene coverage drawn from UW data.

Table 2

Overall PGRNseq performance (HapMap96).

Plex Level	Avg. # Reads (M)	Avg. Unique Aligned Gb	Avg. Mean Quality Score	Avg. % Q30 Bases	Avg. % Targets Hit	Avg. Coverage	Mean % target at > 20x	Mean % targets at > 40x
24	16.7	1.37	36.7	92.1	94.7	496x	94.8	93.4

PGRNseq summary statistics drawn from the BCM HapMap96 data.

Table 3

PGRNseq per-individual concordance vs. orthogonal datasets.

Dataset	Concordance, coverage 10× (mean # overlap)	Concordance, coverage 20× (mean # overlap)	Concordance, coverage 30× (mean # overlap)	Concordance, coverage 50× (mean # overlap)
YRI Deep Trio (n=3)	99.9% (650)	99.9% (547)	100% (354)	100% (182)
CEU Deep Trio (n=3)	99.8% (554)	100% (497)	100% (337)	100% (137)
HapMap 3.3 (n=96)	99.4% (296)	99.4% (296)	99.4% (296)	99.4% (296)
EGP exomes (n=54)	100% (147)	100% (138)	100% (127)	100% (107)

Sample sizes indicate number of overlapping samples between datasets. Concordance calculated for variant sites only with coverage at or above the thresholds listed in the column headers. Table values in parentheses indicate mean number of overlapping variants per individual from which the final percentage was derived.