Title
Exploring the Use of Electronic Health Record-Linked Biorepositories for Pharmacogenomic Application and Discovery

Author
Gonzaludo, Nina

Publication Date
2015

Peer reviewed|Thesis/dissertation

Exploring the Use of Electronic Health Record-Linked Biorepositories for Pharmacogenomic Application and Discovery

by

Nina Gonzaludo

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the GRADUATE

DIVISION of the

This work represents the culmination of 5 years of studying, research, and countless learning experiences at UCSF and beyond.

It is dedicated to my support network: my parents, brother, friends, UCSF family, and all those who inspire, push, and motivate me.

ACKNOWLEDGEMENTS

First, I thank Dr. Pui-Yan Kwok, my mentor and advisor throughout the past few years. From the start, Pui has been a strong supporter of allowing me to shape my own research and pursue whatever interests me, even if it diverged from the primary interests of the lab. With his support, I gained access to the most cutting-edge technologies and projects at UCSF, and I cannot thank him enough for providing me with the resources and opportunities to grow as a scientist.

I also thank Drs. Neil Risch and Rahul Deo, members of my dissertation committee who provided valuable feedback and guidance throughout my research. I acknowledge my oral qualification committee, consisting of Drs. Deanna Kroetz, Kathleen Giacomini, Mark Segal, and Lindsey Criswell, for valuable discussions that helped shape my dissertation work.

I thank all members of the Kwok lab, past and present, for long discussions, brainstorming, problem-solving, criticisms, and praises. I am honored to have worked next to, learned from, and traveled with such a talented group of scientists. I especially thank Paul Tang, Ernest Lam, Blanca Herrera, Angel Mak, Catherine Chu, Annie Poon, and Chin Lin.

I spent a significant amount of my graduate career commuting across the Bay to Kaiser Permanente Division of Research in Oakland. This dissertation work could not have been completed without Dr. Catherine Schaefer, who I thank for helping to facilitate access to this resource, as well as providing guidance and support. I also thank Dilrini Ranatunga and Thomas Hoffman for going above and beyond to answer my questions, Julia Kay for database support, and Diane Oliver for helping to get me on board. Additionally, I thank Ling Shen, Eric Jorgenson, Stan Sciortino, Dana Ludwig, Larry Walter, and Judith Millar.

I especially thank my friends, who helped pushed me through difficult times, celebrated my accomplishments with me, and never gave up hope that I would “hurry up and graduate”. A special thanks to my PSPG classmates; I am privileged to have gone through this experience with such a talented, entertaining, and supportive group of people.

Finally, I am indebted to my mother, father, and brother. Despite the distance and not always understanding my research, my family has been an unwavering source of support and motivation, for which I am extremely grateful.

ABSTRACT

Drug response is well documented to vary considerably among patient groups and populations, as well as within individual patients. Since drug prescribing is often based on population averages of drug response, many patients will not respond, and up to one-third may experience harmful toxicity. Genetics plays a large role in explaining the variability observed in response to different drugs and is an important factor driving precision medicine initiatives. Pharmacogenetic information can be useful in optimizing patient therapy, potentially reducing the cost of hospitalizations and treatment of adverse drug events.

As part of the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH), we analyzed 102,979 members of the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort with genetic information available, along with almost two decades of electronic health record (EHR) data, prescription records, and lifestyle survey results. In one of the largest, most ethnically diverse pharmacogene characterization studies to date, we assessed cohort metabolizer status phenotypes for 7 drug-gene interactions (DGIs) for which there is moderate to strong evidence suggesting the use of pharmacogenetic information to guide therapy. 89% of the cohort had at least one actionable allele for the 7 DGIs in this study, and we observed large variations among ethnicities. Additionally, 17,747 individuals had been prescribed a drug for which they had an actionable or high-risk metabolizer status phenotype. For these individuals, the availability of pharmacogenetic information at point-of-care may have potentially led to a more personalized drug or dosing regimen.

Following this study, we assessed the utility of this resource for deriving two drug response phenotypes: weight gain induced by atypical antipsychotic use and major adverse cardiovascular events in clopiodgrel non-responders. Despite challenges in deriving phenotypes

from the EHR, we were able to extract phenotypes that reflected observed estimates from previously published studies. Using these phenotypes, we performed candidate gene and genome-wide association studies to identify genetic variants associated with response. Altogether, this dissertation demonstrates the potential utility and clinical impact of integrating genetic data with EHRs for pharmacogenetic application and discovery, and provides the foundation for future studies in precision medicine.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION TO CLINICAL PHARMACOGENETICS

1.1 Introduction

Drug response can vary considerably due to numerous factors including genetics, ethnicity, gender, age, or disease status. While intrinsic and environmental factors may affect drug response, drug prescribing is often based on population averages, and in many cases, "one size" does not fit all. In fact, up to one-third of patients may experience harmful toxicity[1], while another portion of patients may not respond at all to a drug.

Since the coining of the term "pharmacogenetics" in the 1950s[2], significant progress has been made in our understanding of how DNA sequence may influence response to various drugs. While research in this area continues at a fast pace, largely helped by advances in technology, translating this information into the clinic remains a challenge. Ultimately, researchers and clinicians are striving towards a future whereby before a patient undergoes drug therapy, their physician has immediate access to the patient's genetic information, and can use it to choose the optimal drug and dosage to best benefit the patient.

The term "precision medicine" was recently coined to represent the idea of tailoring of medical treatment to the individual characteristics of each patient[3]. Numerous efforts, such as the national Precision Medicine Initiative launched by US President Barack Obama in 2015[4], recognize the need for a more comprehensive view of each patient when making treatment decisions. Key to precision medicine is the clinical implementation of pharmacogenomics, which faces numerous challenges, ranging from the development of a well-replicated knowledge base linking genetic variants to drug response, to clinical decision support systems that can incorporate genetic information and help guide treatment. This chapter will review progress that has been made in addressing the multiple barriers of widespread implementation of

pharmacogenomics in clinical practice, and will highlight key examples that have contributed to this field.

1.2 Background

1.2.1 History of Pharmacogenetics

Variation in response to drugs was observed as early as the early 1900s. In 1908, physiologist Sir Archibald Garrod described the notion of "chemical individuality," noting that "every active drug is a poison, when taken in large enough doses; and in some subjects a dose which is innocuous to the majority of people has toxic effects, whereas others show exceptional tolerance of the same drug[5]." Decades later, in the 1950s, significant progress was made in studying genetically determined variations in drug response, with the term "pharmacogenetics" being coined by Vogel in 1959[2].

Early studies were based on clinical observations that patients showed a wide range of plasma or urinary drug concentrations, paired to observations that such variation was often inherited. Initial studies include observations of prolonged apnea after taking succinylcholine due to a deficiency in pseudocholinesterase[6] (an autosomal recessive trait[7]), and a polymorphism in *N*-acetyltransferase-2 (*NAT2*) that causes deficient biotransformation of isoniazid, a treatment for tuberculosis, resulting in toxicity[8,9]. Studies of the antimalarial drug primaquine given in the South Pacific during World War II showed a higher rate of hemolytic anemia among African Americans with glucose-6 phosphate dehydrogenase (*G6PD*) deficiency[10-12], hinting towards the role of ancestry in clinically observed traits.

1.2.2 Mechanisms of Variability of Drug Concentration and Effect

In the 1970s, two independent pharmacokinetic studies of the antihypertensive drug debrisoquine[13] and the antiarrhythmic sparteine[14], observed that 5 to 10 percent of subjects

displayed low urinary concentration of the metabolite and higher plasma concentrations of the parent drug, along with exaggerated drug response. This effect was determined to be monogenic and inherited as an autosomal recessive trait[15]. It was later found that deficiency in the same cytochrome-P450 (CYP450) enzyme was responsible for the observations seen with both drugs[16], and in 1988, the specific enzyme, *CYP2D6*, was cloned[17].

The CYP450 superfamily of enzymes are the most important enzymes for catalyzing phase I (e.g. oxidation, reduction, and hydrolysis) drug metabolism reactions[18]. Both phase I and phase II (conjugation reactions such as acetylation, glucoronidation, sulfation, and methylation) reactions generally convert drugs to metabolites that are more water-soluble and easier to excrete[19]. *CYP2D6*, one of the first CYP450 enzymes to be cloned, remains one of the most well studied genes involved in drug metabolism and is involved in the bioactivation of 20-25% of marketed drugs[20]. Variants in CYP450 genes that result in protein changes can generate a range of responses to substrates of each enzyme. Given the highly polymorphic nature of *CYP2D6*, a spectrum of drug responders, ranging from poor metabolizers ("PMs", with low levels of metabolite formed compared to parent drug) to ultra-rapid metabolizers ("UMs", who rapidly convert parent drug to metabolite) are observed for *CYP2D6* substrates[21]. Variants in other, less polymorphic CYP450 genes, such as *CYP2C9* and *CYP2C19*, may result in a less variable, yet still clinically relevant, range of responses. Notably, frequency rates of functionally-relevant CYP450 variants often vary greatly by ancestry, resulting in different rates of PMs in different populations. The US Food and Drug Administration (FDA) first approved a laboratory pharmacogenetic test in 2005: the AmpliChip® CYP450 Test (Roche Molecular Diagnostics, NJ, USA), which utilized Affymetrix (CA, USA) microarray technology to genotype *CYP2D6* and *CYP2C19* alleles[22]. Since then, commercially available pharmacogenetic tests have been

developed for the majority of the pharmacogenetic biomarkers that the FDA has included in labels for over 100 drugs[23].

Candidate gene pharmacogenetic studies also considered other mechanisms underlying variability in drug concentration and effect, such as variation in drug uptake, distribution, and elimination pathways, as well as drug targets. Research on the latter has led to significant developments in targeted therapy. Variants in membrane-bound efflux transporters, such as the ATP-binding cassette transporter family (ABCs) and organic anion transporters (OATs and OATPs), have also been shown to modulate responses to a variety of drugs[24].

1.2.3 Technology and Approaches

Early pharmacogenetic studies mainly identified monogenic traits with relatively common genetic polymorphisms[19]. Since these studies, significant improvements in both cost and efficiency of genotyping and next-generation sequencing technologies, computing, and analysis methods have advanced our understanding of genetic variation influencing drug response. Most pharmacogenetic studies to date focused on the DNA sequence of candidate genes or pathways involved in drug response. The availability of tools that enable querying variants genome-wide allow for more unbiased approaches, contributing to the development of "pharmacogenomic" research. Additionally, new technologies allow for the investigation of variation beyond DNA sequence, enabling researchers to study differences in RNA expression, methylation sites, or the microbiome, for example. At the same time, genetic research has shifted focus towards characterization of more complex, polygenic traits, including common diseases and drug response.

1.2.4 Clinical Pharmacogenetics Examples

As technology improves, so too does the ability to both identify and validate key genes involved in drug response in clinical settings. Table 1.1 highlights well-studied clinical pharmacogenetic examples. Given the number of genetic studies investigating these examples in multiple populations, such drugs are most likely to be among the first candidates for implementation of personalized medicine in clinical practice. The Clinical Pharmacogenetics Implementation Consortium (CPIC) of the National Institute of Health's Pharmacogenomics Research Network was created in part to provide guidance on clinical interpretation of genetic test results for certain gene/drug pairs based on peer-reviewed evidence in literature[25]. In addition to the drugs listed in Table 1.1, the FDA has amended drug labels for over 100 drugs to include information about genetic variants that may influence drug response[23]. Pharmacogenetic guidance may be particularly useful when administering high-risk drugs with narrow therapeutic indices (NTI) or only one major pathway of metabolism or elimination[26]. In such instances, variants leading to deficiencies in a single gene may significantly alter drug concentration and response. For NTI drugs, this may result in toxicity and adverse events. Such high-risk drugs are also more susceptible to drug-drug interactions, in which another administered drug may compete with the single metabolism or elimination pathway. Chapter 2 describes a characterization study of a panel of pharmacogenes that are included in several CPIC guidelines for utilizing genetics when prescribing certain drugs.

Table 1.1. Examples of drugs with genetic variants influencing clinical response

Drug	*Gene(s)*	*Clinical Effect*	*Year*	*Reference(s)*
6-mercaptopurine	TPMT	Decreased dose requirements	1980	[27]
codeine	CYP2D6	Increased risk of toxicity in UMs, insufficient response in PMs	1991	[28]
warfarin	CYP2C9, VKORC1	Decreased dose requirements, possible increased risk for bleeding	1994	[29-31]
irinotican	UGT1A1	Increased risk for adverse events	1998	[32]
abacavir	HLA	Skin reaction	2002	[33,34]
carbamazepine	HLA	Skin reaction, Stevens-Johnson Syndrome	2004	[35]
tamoxifen	CYP2D6	Increased risk for relapse, poor outcomes in IMs and PMs	2005	[36,37]
simvastatin	SLCO1B1	Increased risk for myopathy	2008	[38]
clopidogrel	CYP2C19, ABCB1	Decreased effect, increased risk for poor outcomes in PMs	2009	[39-41]

Cancer and cardiovascular disease therapeutics have been particularly impacted by pharmacogenetic discoveries. Mercaptopurine, an immunosuppressant chemotherapy agent used to treat acute lymphoblastic leukemia and other neoplasias, is an early example pharmacogenetic testing in clinical practice[19]. Thiopurine S-methyltransferase (*TPMT*) was found to metabolize mercaptopurine[27], and studies have observed that patients with a single common allele in *TPMT* (TPMT*3A allele) are at elevated risk for life-threatening myelosuppresion[42]. Irinotican, a topoisomerase 1 inhibitor used to treat colon and rectal cancers, is converted to its active metabolite by hydrolysis, which is then inactivated by UDP glucoronosyltransferase 1 family,

polypeptide A1 (*UGT1A1*)[32]. Deficiency in *UGT1A1* due to genetic variants has been associated with irinotecan toxicity and excessive bilirubin[43,44]. Tamoxifen, widely used to treat estrogen receptor positive breast cancer, undergoes multiple pathways of metabolism by a number of different CYP450s[45]. *CYP2D6* plays a major role in the formation of the potent metabolite endoxofin[46], and PMs have been found to have an ineffective response to tamoxifen, resulting in relapse or poor outcomes[36,37].

In cardiovascular disease therapy, warfarin remains the most extensively studied pharmacogenetic example. Warfarin, a widely prescribed anticoagulant, has a dosing range that varies by a factor of 10- to 20-fold among patients[47,48], with incorrect dosing leading to severe adverse events. So far, the International Warfarin Pharmacogenetics Consortium has analyzed the relationship between CYP2C9 and VKORC1 genotypes with warfarin dose in over 5000 patients of various ancestries, and has created dosing algorithms that can be used clinically upon prescribing[48]. Clopidogrel, a cardiovascular drug that prevents platelet aggregation and the focus of chapter 4, is metabolized by multiple pharmacogenes, with drug response primarily influenced by *CYP2C19* genotype[49].

1.2.5 Electronic Health Records and Biorepositories

While researchers continue contributing to a growing body of knowledge linking genetic variants to drug response, multiple barriers still exist for integrating such information into clinical practice. One major barrier is infrastructure to collect, store, and present pharmacogenetic information to clinicians. Setting up such infrastructure in countries such as the United States, with multiple disparate health care systems, will continue to remain a challenge. Nonetheless, investigators are looking into implementing pharmacogenetics within their own institutions, with many institutions sharing best practices and early results[50-53]. Adoption of

electronic health record (EHR) systems is a key first step in implementation, and as of 2014, 3 out of 4 non-federal acute care hospitals (76%) have at least a basic EHR system in place, an eight-fold increase since 2008[54]. Such systems allow for integration of clinical decision support rules based on a pharmacogenetic knowledge base. For example, when a physician orders a certain drug, a rule-based engine may cause an alert to pop-up, advising them to order a pharmacogenetic test first[51].

While EHRs are key for implementation, they are also proving to be excellent tools for pharmacogenomic discovery and collection of drug response data. Phenotypes for the majority of published pharmacogenetic studies have largely been obtained from randomized clinical trials (RTCs)[55]. Data collection for such studies has been costly and time-consuming. EHRs rich in longitudinal medical data can potentially be used to quickly extract phenotypic information for use in pharmacogenetic studies, and may capture clinical covariates, pharmacy, and lab information not typically collected in RTCs. Chapters 3 and 4 describe examples of using this type of resource for deriving drug response phenotypes.

While EHRs are a robust source of phenotypic information, biorepositories or biobanks of collected patient tissue samples may serve as a resource for genotypic information. Besides typical costs to maintain a biorepository, genotyping efforts and data generation, analysis, and integration are another challenge. Ethical issues of consent and dissemination of analysis findings are also a major concern, as are issues of educating both patients and physicians on interpretation of pharmacogenetic testing results[56]. Nonetheless, several institutions and consortia, such as the Electronic Medical Records and Genomics (eMERGE) Network[57] and the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH)[58], have

confronted many of these hurdles, and made significant progress in integrating EHR systems with genetic data.

1.3 Statement of Purpose

In this study, we utilize data from the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, part of the Kaiser Permanente RPGEH. The GERA cohort consists of 102,979 Kaiser Permanente Northern California (KPNC) adult members that have undergone genome-wide genotyping using Affymetrix Axiom technology[59,60]. In addition to genetic data, this resource includes extensive pharmacy, lab, procedure, diagnosis, and lifestyle data for cohort members, derived from a comprehensive EHR system and patient surveys. A description of this resource has been previously described[61] and is also available at dbGAP accession number phs000674.v1.p1. In this dissertation, we leverage these data to assess the impact of pharmacogenetic information on clinically observed drug response, as well as utilize EHR-derived drug response phenotypes for genetic studies.

1.4 References

1. Wilke, R. A. & Dolan, M. E. Genetics And Variable Drug Response. *JAMA* **306,** 306–307 (2011).
2. Vogel, F. Moderne Probleme der Humangenetik. *Ergeb Inn Med Kinderheilkd* **12,** 52–125 (1959).
3. Committee on a Framework for Development a New Taxonomy of Disease, Board on Life Sciences, Division on Earth and Life Studies & National Research Council. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease* (National Academies Press, Washington, D.C., USA, 2011).
4. WhiteHouse.gov. Precision Medicine Initiative. https://www.whitehouse.gov/precision-

medicine (2015).

5. Garrod, A. E. *Inborn Errors of Metabolism*. (Henry Frowde and Hodder & Stoughton, London, United Kingdom, 1923).

6. Kalow, W. Familial Incidence Of Low Pseudocholinesterase Level. *Lancet* **268,** 576–577 (1956).

7. Kalow, W. & Staron, N. On Distribution And Inheritance Of Atypical Forms Of Human Serum Cholinesterase, As Indicated By Dibucaine Numbers. *Can J Med Sci* **35,** 1305–1320 (1957).

8. Evans, D. A. P., Manley, K. A. & McKusick, V. A. Genetic Control Of Isoniazid Metabolism In Man. *BMJ* **2,** 485–491 (1960).

9. Vatsis, K. P., Martell, K. J. & Weber, W. W. Diverse Point Mutations In The Human Gene For Polymorphic N-Acetyltransferase. *Proc Natl Acad Sci USA* **88,** 6333–6337 (1991).

10. Clayman, C. B. *et al.* Toxicity Of Primaquine In Caucasians. *JAMA* **149,** 1563–1568 (1952).

11. Alving, A. S., Carson, P. E., Flanagan, C. L. & Ickes, C. E. Enzymatic Deficiency In Primaquine-Sensitive Erythrocytes. *Science* **124,** 484–485 (1956).

12. Beutler, E, Dern, R. J. & Alving, A. S. The Hemolytic Effect Of Primaquine. Vi. An In Vitro Test For Sensitivity Of Erythrocytes To Primaquine. *J Lab Clin Med* **45,** 40–50 (1955).

13. Mahgoub, A., Idle, J. R., Dring, L. G., Lancaster, R. & Smith, R. L. Polymorphic Hydroxylation Of Debrisoquine In Man. *Lancet* **2,** 584–586 (1977).

14. Eichelbaum, M., Spannbrucker, N., Steincke, B. & Dengler, H. J. Defective N-Oxidation

Of Sparteine In Man: A New Pharmacogenetic Defect. *Eur J Clin Pharmacol* **16,** 183–187 (1979).

15. Meyer, U. A. Pharmacogenetics - Five Decades Of Therapeutic Lessons From Genetic Diversity. *Nat Rev Genet* **5,** 669–676 (2004).

16. Bertilsson, L., Dengler, H. J., Eichelbaum, M. & Schulz, H. U. Pharmacogenetic Covariation Of Defective N-Oxidation Of Sparteine And 4-Hydroxylation Of Debrisoquine. *Eur J Clin Pharmacol* **17,** 153–155 (1980).

17. Gonzalez, F. J. *et al.* Characterization Of The Common Genetic Defect In Humans Deficient In Debrisoquine Metabolism. *Nature* **331,** 442–446 (1988).

18. Wilkinson, G. R. in *Goodman And Gilman's The Pharmacological Basis Of Therapeutics* 3–29 (McGraw-Hill Professional, New York, New York, USA, 2001).

19. Weinshilboum, R. Inheritance And Drug Response. *N Engl J Med* **348,** 529–537 (2003).

20. Ingelman-Sundberg, M. Genetic Polymorphisms Of Cytochrome P450 2D6 (CYP2D6): Clinical Consequences, Evolutionary Aspects And Functional Diversity. *Pharmacogenomics J* **5,** 6–13 (2005).

21. Zhou, S.-F. Polymorphism Of Human Cytochrome P450 2D6 And Its Clinical Significance: Part I. *Clin Pharmacokinet* **48,** 689–723 (2009).

22. A, S. *et al.* Realities And Expectations Of Pharmacogenomics And Personalized Medicine: Impact Of Translating Genetic Knowledge Into Clinical Practice. *Pharmacogenomics* **11,** 1149–1167 (2010).

23. U.S. Food and Drug Administration. Table of Pharmacogenomic Biomarkers in Drug Labels. http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.ht

m (2015).

24. Sissung, T. M. T. *et al.* Transporter Pharmacogenetics: Transporter Polymorphisms Affect Normal Physiology, Diseases, And Pharmacotherapy. *Discov Med* **13,** 19–34 (2012).

25. Relling, M. V. & Klein, T. E. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther* **89,** 464–467 (2011).

26. Roden, D. M. D., Wilke, R. A. R., Kroemer, H. K. H. & Stein, C. M. C. Pharmacogenomics: The Genetics Of Variable Drug Responses. *Circulation* **123,** 1661–1670 (2011).

27. Weinshilboum, R. M. & Sladek, S. L. Mercaptopurine Pharmacogenetics: Monogenic Inheritance Of Erythrocyte Thiopurine Methyltransferase Activity. *Am J Hum Genet* **32,** 651–662 (1980).

28. Desmeules, J., Gascon, M. P., Dayer, P. & Magistris, M. Impact Of Environmental And Genetic Factors On Codeine Analgesia. *Eur J Clin Pharmacol* **41,** 23–26 (1991).

29. Rettie, A. E., Wienkers, L. C., Gonzalez, F. J., Trager, W. F. & Korzekwa, K. R. Impaired (S)-Warfarin Metabolism Catalysed By The R144C Allelic Variant Of CYP2C9. *Pharmacogenet Genomics* **4,** 39 (1994).

30. Aithal, G. P., Day, C. P., Kesteven, P. J. & Daly, A. K. Association Of Polymorphisms In The Cytochrome P450 CYP2C9 With Warfarin Dose Requirement And Risk Of Bleeding Complications. *Lancet* **353,** 717–719 (1999).

31. Rieder, M. J. *et al.* Effect Of VKORC1 Haplotypes On Transcriptional Regulation And Warfarin Dose. *N Engl J Med* **352,** 2285–2293 (2005).

32. Iyer, L. *et al.* Genetic Predisposition To The Metabolism Of Irinotecan (CPT-11). Role Of

Uridine Diphosphate Glucuronosyltransferase Isoform 1A1 In The Glucuronidation Of Its Active Metabolite (SN-38) In Human Liver Microsomes. *J Clin Invest* **101,** 847–854 (1998).

33. Mallal, S. *et al.* Association Between Presence Of HLA-B*5701, HLA-DR7, And HLA-DQ3 And Hypersensitivity To HIV-1 Reverse-Transcriptase Inhibitor Abacavir. *Lancet* **359,** 727–732 (2002).

34. Mallal, S. *et al.* HLA-B*5701 Screening for Hypersensitivity to Abacavir. *N Engl J Med* **358,** 568–579 (2008).

35. Chung, W.-H. *et al.* Medical Genetics: A Marker For Stevens–Johnson Syndrome. *Nature* **428,** 486–486 (2004).

36. Goetz, M. P. *et al.* Pharmacogenetics Of Tamoxifen Biotransformation Is Associated With Clinical Outcomes Of Efficacy And Hot Flashes. *J Clin Oncol* **23,** 9312–9318 (2005).

37. Jin, Y. *et al*. CYP2D6 Genotype, Antidepressant Use, And Tamoxifen Metabolism During Adjuvant Breast Cancer Treatment. *J Natl Cancer Inst* **97**, 30-39 (2005).

38. SEARCH Collaborative Group *et al.* SLCO1B1 Variants And Statin-Induced Myopathy--A Genomewide Study. *N Engl J Med* **359,** 789–799 (2008).

39. Mega, J., Close, S., Wiviott, S. & Shen, L. Cytochrome P-450 Polymorphisms and Response to Clopidogrel. *N Engl J Med* **360,** 354–362 (2009).

40. Collet, J.-P. *et al.* Cytochrome P450 2C19 Polymorphism In Young Patients Treated With Clopidogrel After Myocardial Infarction: A Cohort Study. *Lancet* **373,** 309–317 (2009).

41. Simon, T. *et al.* Genetic Determinants of Response to Clopidogrel and Cardiovascular Events. *N Engl J Med* **360,** 363–375 (2009).

42. Lennard, L., Van Loon, J. A. & Weinshilboum, R. M. Pharmacogenetics Of Acute

Azathioprine Toxicity: Relationship To Thiopurine Methyltransferase Genetic Polymorphism. *Clin Pharmacol Ther* **46,** 149–154 (1989).

43. Iyer, L. *et al.* Phenotype-Genotype Correlation Of In Vitro SN-38 (Active Metabolite Of Irinotecan) And Bilirubin Glucuronidation In Human Liver Tissue With UGT1A1 Promoter Polymorphism. *Clin Pharmacol Ther* **65,** 576-582 (1999).

44. Marcuello, E. *et al.* UGT1A1 Gene Variations And Irinotecan Treatment In Patients With Metastatic Colorectal Cancer. *Br J Cancer* **91,** 678–682 (2004).

45. Crewe, H. K., Ellis, S. W., Lennard, M. S. & Tucker, G. T. Variable Contribution Of Cytochromes P450 2D6, 2C9 And 3A4 To The 4-Hydroxylation Of Tamoxifen By Human Liver Microsomes. *Biochem Pharmacol* **53,** 171–178 (1997).

46. Lee, K.-H., Ward, B. A., Desta, Z., Flockhart, D. A. & Jones, D. R. Quantification Of Tamoxifen And Three Metabolites In Plasma By High-Performance Liquid Chromatography With Fluorescence Detection: Application To A Clinical Trial. *J Chromatogr B Analyt Technol Biomed Life Sci* **791,** 245–253 (2003).

47. Takeuchi, F. *et al.* A Genome-Wide Association Study Confirms VKORC1, CYP2C9, and CYP4F2 as Principal Genetic Determinants of Warfarin Dose. *PLoS Genet* **5,** e1000433 (2009).

48. International Warfarin Pharmacogenetics Consortium *et al.* Estimation Of The Warfarin Dose With Clinical And Pharmacogenetic Data. *N Engl J Med* **360,** 753–764 (2009).

49. Ahmad, T., Voora, D. & Becker, R. C. The Pharmacogenetics Of Antiplatelet Agents: Towards Personalized Therapy? *Nat Rev Cardiol* **8,** 560–571 (2011).

50. Bielinski, S. J. *et al.* Preemptive Genotyping For Personalized Medicine: Design Of The Right Drug, Right Dose, Right Time-Using Genomic Data To Individualize Treatment

Protocol. *Mayo Clin. Proc.* **89,** 25–33 (2014).

51. Bell, G. C. *et al.* Development And Use Of Active Clinical Decision Support For Preemptive Pharmacogenomics. *J Am Med Inform Assoc* **21**, e93-99 (2013). doi:10.1136/amiajnl-2013-001993

52. Kho, A. N. *et al*. Electronic Medical Records For Genetic Research: Results Of The Emerge Consortium. *Sci Transl Med* **3**, 79re1 (2011). doi: 10.1126/scitranslmed.3001807

53. Pulley, J. M. J. *et al.* Operational Implementation Of Prospective Genotyping For Personalized Medicine: The Design Of The Vanderbilt PREDICT Project. *Clin Pharmacol Ther* **92,** 87–95 (2012).

54. The Office of the National Coordinator for Health Information Technology. ONC Data Brief No. 23: Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2014. http://healthit.gov/sites/default/files/data-brief/2014HospitalAdoptionDataBrief.pdf (2015).

55. Fefer, P. & Matetzky, S. The Genetic Basis Of Platelet Responsiveness To Clopidogrel. A Critical Review Of The Literature. *Thromb Haemost* **106,** 203–210 (2011).

56. McKinnon, R. A., Ward, M. B. & Sorich, M. J. A Critical Analysis Of Barriers To The Clinical Implementation Of Pharmacogenomics. *Ther Clin Risk Manag* **3,** 751–759 (2007).

57. eMERGE. The eMERGE Network: Electronic Medical Records and Genomes. http://emerge.mc.vanderbilt.edu (2014).

58. Kaiser Permanente Division of Research. The Research Program on Genes, Environment, & Health. http://www.rpgeh.kaiser.org (2015).

59. Hoffmann, T. J. *et al.* Design And Coverage Of High Throughput Genotyping Arrays

Optimized For Individuals Of East Asian, African American, And Latino Race/Ethnicity Using Imputation And A Novel Hybrid SNP Selection Algorithm. *Genomics* **98,** 422–430 (2011).

60. Hoffmann, T. J. *et al.* Next Generation Genome-Wide Association Tool: Design And Coverage Of A High-Throughput European-Optimized SNP Array. *Genomics* **98,** 79–89 (2011).

61. Kvale, M. N. *et al.* Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200,** 1051–1060 (2015).

2 PHARMACOGENE CHARACTERIZTAION

2.1 Introduction

Incorporation of pharmacogenetic information into routine clinical practice, a necessary step of precision medicine efforts, faces many hurdles[1,2]. One such hurdle is the timely availability of accurate genetic information to clinicians when it is needed, at point of care. Generally, lab values, symptoms, or a certain diagnosis may spur physicians to order pharmacogenetic testing prior to prescribing a drug, but there can be significant delays in the return of actionable information[2]. Current approaches of obtaining such information can also be costly and impractical[3].

While genetic information can be useful for guiding personalized therapy, there are multiple challenges in translating genotypic data into clinically actionable information. For instance, researchers and clinicians use different nomenclatures to identify pharmacogenetic variants[4]. Genetic data in the form of single nucleotide polymorphisms are typically annotated with a dbSNP identifier in research studies. Pharmacogenetic information, which can be found in boxed warnings on the package insert of certain drugs, is often communicated to physicians in the form of star allele nomenclature[4]. This nomenclature system was initially created to annotate variants in cytochrome p450 enzymes[5], but has been adopted for the annotation of other pharmacogenes. In this system, numbered alleles represent functional variants within a gene. Star allele status forms the basis of common pharmacogenetic metabolizer phenotypes, such as extensive metabolizers (EMs) who benefit from the standard recommended drug regimen, or poor metabolizers (PMs) who are unable to metabolize a drug and are thus at risk of adverse drug reactions or a lack of therapeutic effect. Translating between the two nomenclatures is a

non-trivial but necessary initial task[6] that allows for characterization of pharmacogenes and an assessment of the frequency of drug gene interactions with implications for clinical use.

The Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) was created to enable research on genetic and environmental factors that affect health and disease by combining several data sources, including electronic health records (EHR) and genetic data derived from collected biospecimens. The Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort includes over 100,000 adult members of Kaiser Permanente Northern California (KPNC) who are participants of RPGEH and have undergone genome-wide genotyping using Affymetrix Axiom technology, as previously described[7]. Given that the median age of cohort members was 64 years old at recruitment, and polypharmacy and drug use generally increase with age[8], this resource presents the opportunity to quantify decades of drug response data in combination with genetic data in a large, diverse population.

In this chapter we utilize a diverse, real-world dataset to assess the variability in drug response phenotypes across ethnicities and patient groups. This chapter highlights the ethnicity-specific differences in frequency of clinically actionable pharmacogenes for 7 drug-gene interactions (DGIs). We utilized guidelines from the Clinical Pharmacogenetics Implementation Consortium (CPIC) [9] to determine which pharmacogenetic alleles to consider actionable. We also address clinical relevance by assessing retrospective medication exposure to 34 drugs with moderate to strong clinical evidence supporting the use of genetic information for prescribing, and tie exposure back to pharmacogene status for the 7 drugs in this study. This results in the largest, most diverse pharmacogene characterization study to date, highlighting the potential impact of preemptive genotyping, particularly among different ethnicities.

2.2 Results

2.2.1 Study Population

102,979 RPGEH members were successfully genotyped, representing the GERA cohort (Table 2.1). Eighty-one percent of cohort members are Non-Hispanic White, while 7.8%, 7.0%, 3.3% are Asian, Latino, and African American, respectively. 0.74% was classified as other or uncertain. 58% of cohort members are female. Overall, at sample collection, cohort members had a median age of 64 years and 14.9 years of prescription records available as part of this resource. 75% of the cohort had at least a decade of prescription records available.

Table 2.1 GERA Cohort Demographics

	All GERA	White	African American	Latino	Asian	Other/ Uncertain
Total (%)	102,979	83,513 (81%)	3,380 (3.3%)	7,246 (7.0%)	8,082 (7.8%)	758 (0.073%)
% Male	42	42	42	38	43	47
Median Age at Sample Collection in years (interquartile range)	64 (55-73)	65 (56-73)	62 (52-70)	58 (47-68)	59 (48-68)	64 (56-74)
Median EHR Rx Exposure Time in years (interquartile range)	14.9 (10.7-17.1)	15.0 (10.8-17.1)	15.0 (11.5-17.1)	14.5 (10.3-17.0)	14.5 (10.1-17.1)	15.1 (9.9-17.1)

EHR Rx exposure time indicates length of prescription record available for cohort members. See Materials & Methods – Study Population.

2.2.2 Pharmacogene Star Alleles

Pharmacogenes of interest were selected according to CPIC guideline availability (Table 2.2) with strong clinical evidence of genetic association, as well as presence of prescriptions in the GERA cohort, as described in Materials and Methods. 7 DGIs were chosen, with prescription

frequencies shown in Figure 2.1. Star allele frequencies for the pharmacogenes in the 7 DGIs were calculated for each ethnic group and are shown in Table 2.3. Observed star allele frequencies were generally comparable to published estimates, although reports are limited for some ethnicities [10].

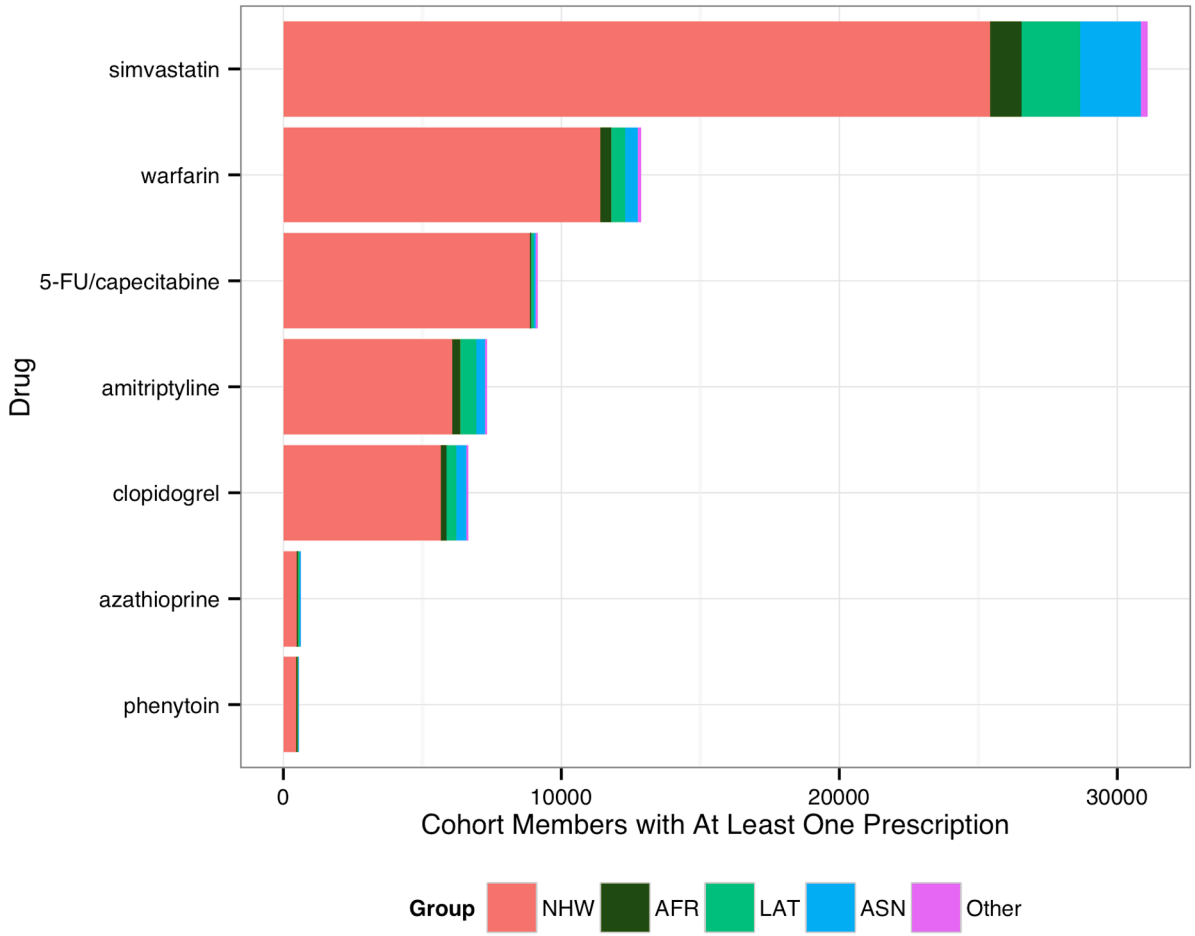
Table 2.2 Summary of actionable CPIC guidelines for selected drug-gene interactions

Drug(s)	Gene(s)	Diplotype	Metabolizer Phenotype	Recommended Action
clopidogrel	CYP2C19	*1/*2	IM	Consider alternative antiplatelet agent (e.g. prasugrel, ticagrelor)[11]
		*1/*3	IM	
		*2/*17	IM	
		*2/*2	PM	
		*2/*3	PM	
		*3/*3	PM	
simvastatin	SLCO1B1	*1/*5	IM	Consider a lower dose; if suboptimal efficacy, consider an alternative statin[12]
		*5/*5	PM	Prescribe a lower dose or consider an alternative statin; consider routine CK surveillance[12]
azathioprine	TPMT	1 nonfunctional allele (*2, *3A, *3B, *3C, or *4)	IM	Consider starting at 30-70% of target dose and titrate based on tolerance. Allow 2-4 weeks to reach steady state after each dose adjustment[13]
		2 nonfunctional alleles (*2, *3A, *3B, *3C, or *4)	PM	Consider alternative agents. For azathioprine, start with drastically reduced doses (reduce daily dose by 10-fold and dose thrice weekly instead of daily) and adjust doses of azathioprine based on degree of myelosuppression and disease-specific guidelines. Allow 4-6 weeks to reach steady state after each dose adjustment[13]
warfarin	CYP2C9, VKORC1	*1/*1, *2/*2	IM	3-4 mg/day (from 5-7 mg/day standard dose)[14]
		*1/*2, *1/*2	IM	

		*1/*2, *2/*2	IM	
		*1/*3, *1/*1	IM	
		*1/*3, *1/*2	IM	
		*1/*3, *2/*2	PM	0.5-2 mg/day (from 5-7 mg/day standard dose)[14]
		*2/*2, *1/*1	IM	3-4 mg/day (from 5-7 mg/day standard dose)[14]
		*2/*2, *1/*2	IM	
		*2/*2, *2/*2	PM	0.5-2 mg/day (from 5-7 mg/day standard dose)[14]
		*2/*3, *1/*1	IM	3-4 mg/day (from 5-7 mg/day standard dose)[14]
		*2/*3, *1/*2	PM	0.5-2 mg/day (from 5-7 mg/day standard dose)[14]
		*2/*3, *2/*2	PM	
		*3/*3, any	PM	
amitriptyline	CYP2C19	*1/*17	UM	Consider alternative drug not metabolized by CYP2C19, or use TDM[15]
		*17/*17	UM	
		*2/*2	PM	Consider a 50% reduction of recommended starting dose, use TDM[15]
		*2/*3	PM	
		*3/*3	PM	
5-fluorouracil, capecitabine	DPYD	1 nonfunctional allele (*2A, *13, or rs67376798)	IM	Start with at least a 50% reduction in starting dose, followed by titration of dose based on toxicity or pharmacokinetic test (if available)[16]
		2 nonfunctional alleles (*2A, *13, or rs67376798)	PM	Select alternative drug[16]
phenytoin	CYP2C9	1 nonfunctional allele (*2 or *3)	EM	Consider 25% dose reduction or alternate drug (if HLA-B*15:02). Use TDM to guide dose adjustments[17]
		2 nonfunctional alleles (*2 or *3)	IM	Consider 50% dose reduction or alternate drug (if HLA-B*15:02). Use TDM to guide dose adjustments[17]

Table does not include metabolizer phenotypes and star alleles that do not warrant a change in therapy. All drug-gene pairs have moderate to strong clinical evidence in favor of changing prescribing based on genetic information[9]. A cohort member with a star allele diplotype listed in this table is considered to have an actionable phenotype for the relevant drug-gene pair. UM = ultra-rapid metabolizer, EM = extensive metabolizer, IM = intermediate metabolizer, PM = poor metabolizer, TDM = therapeutic drug monitoring.

Figure 2.1 Medication exposure by ethnicity



Frequency of cohort members with at least one outpatient record prescription for each of the study drugs, colored by ethnicity. NHW = Non-Hispanic White, AFR = African American, LAT = Latino, ASN = Asian.

Table 2.3 Actionable star allele frequencies

		All GERA	NHW	AFR	LAT	ASN	Other
		102979	83513	3380	7246	8082	758
CYP2C19							
*1/*2		26819 (26)	20651 (24.7)	977 (28.9)	1570 (21.7)	3417 (42.3)	204 (26.9)
*2/*2		2798 (2.7)	1814 (2.2)	101 (3)	106 (1.5)	762 (9.4)	15 (2)
*1/*3		998 (1)	28 (0)	12 (0.4)	21 (0.3)	933 (11.5)	4 (0.5)
*3/*3		14 (0)	0 (0)	0 (0)	0 (0)	14 (0.2)	0 (0)
*2/*3		313 (0.3)	3 (0)	1 (0)	3 (0)	306 (3.8)	0 (0)
*1/*17		32615 (31.7)	28902 (34.6)	1156 (34.2)	1822 (25.1)	501 (6.2)	234 (30.9)
*17/*17		4807 (4.7)	4390 (5.3)	181 (5.4)	177 (2.4)	27 (0.3)	32 (4.2)
*2/*17		5849 (5.7)	5118 (6.1)	264 (7.8)	270 (3.7)	150 (1.9)	47 (6.2)
SLCO1B1							
*1/*5		25833 (25.1)	22022 (26.4)	297 (8.8)	1580 (21.8)	1734 (21.5)	200 (26.4)
*5/*5		2331 (2.3)	2038 (2.4)	14 (0.4)	113 (1.6)	153 (1.9)	13 (1.7)
TPMT							
*1/*2		472 (0.5)	441 (0.5)	0 (0)	25 (0.3)	5 (0.1)	1 (0.1)
*2/*2		6 (0)	3 (0)	1 (0)	2 (0)	0 (0)	0 (0)
*1/*3A		7165 (7)	6359 (7.6)	102 (3)	555 (7.7)	88 (1.1)	61 (8)
*3A/*3A		162 (0.2)	145 (0.2)	3 (0.1)	12 (0.2)	0 (0)	2 (0.3)
*1/*3B		7183 (7)	6374 (7.6)	104 (3.1)	556 (7.7)	88 (1.1)	61 (8)
*3B/*3B		162 (0.2)	145 (0.2)	3 (0.1)	12 (0.2)	0 (0)	2 (0.3)
*1/*3C		8222 (8)	6908 (8.3)	321 (9.5)	655 (9)	268 (3.3)	70 (9.2)
*3C/*3C		207 (0.2)	180 (0.2)	7 (0.2)	14 (0.2)	4 (0)	2 (0.3)
*2/*3A		27 (0)	26 (0)	0 (0)	1 (0)	0 (0)	0 (0)
*2/*3B		27 (0)	26 (0)	0 (0)	1 (0)	0 (0)	0 (0)
*2/*3C		30 (0)	28 (0)	0 (0)	1 (0)	1 (0)	0 (0)
*3A/*3B		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
*3A/*3C		0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
*3B/*3C		5 (0)	4 (0)	0 (0)	1 (0)	0 (0)	0 (0)
CYP2C9	**VKORC1**						

*1/*2	-	20826 (20.2)	18995 (22.7)	236 (7)	1260 (17.4)	173 (2.1)	162 (21.4)
*2/*2	-	1584 (1.5)	1490 (1.8)	6 (0.2)	73 (1)	5 (0.1)	10 (1.3)
*1/*2	*1/*2	9855 (9.6)	9005 (10.8)	57 (1.7)	638 (8.8)	76 (0.9)	79 (10.4)
*1/*2	*2/*2	3265 (3.2)	2963 (3.5)	14 (0.4)	222 (3.1)	42 (0.5)	24 (3.2)
*2/*2	*1/*2	752 (0.7)	710 (0.9)	1 (0)	36 (0.5)	1 (0)	4 (0.5)
*2/*2	*2/*2	244 (0.2)	237 (0.3)	0 (0)	7 (0.1)	0 (0)	0 (0)
*1/*3	-	12214 (11.9)	10743 (12.9)	122 (3.6)	713 (9.8)	549 (6.8)	87 (11.5)
*3/*3	-	334 (0.3)	310 (0.4)	0 (0)	12 (0.2)	12 (0.1)	0 (0)
*1/*3	*1/*2	5728 (5.6)	5136 (6.2)	39 (1.2)	371 (5.1)	147 (1.8)	35 (4.6)
*1/*3	*2/*2	2072 (2)	1619 (1.9)	6 (0.2)	121 (1.7)	314 (3.9)	12 (1.6)
*3/*3	*1/*2	158 (0.2)	148 (0.2)	0 (0)	6 (0.1)	4 (0)	0 (0)
*3/*3	*2/*2	53 (0.1)	51 (0.1)	0 (0)	0 (0)	2 (0)	0 (0)
*2/*3	-	1609 (1.6)	1493 (1.8)	4 (0.1)	93 (1.3)	12 (0.1)	7 (0.9)
*2/*3	*1/*2	762 (0.7)	711 (0.9)	2 (0.1)	43 (0.6)	4 (0)	2 (0.3)
*2/*3	*2/*2	257 (0.2)	233 (0.3)	1 (0)	20 (0.3)	2 (0)	1 (0.1)
-	*1/*2	46328 (45)	39539 (47.3)	807 (23.9)	3568 (49.2)	2046 (25.3)	368 (48.5)
-	*2/*2	19756 (19.2)	12719 (15.2)	76 (2.2)	1358 (18.7)	5484 (67.9)	119 (15.7)
DPYD							
*1/*2A		221 (0.2)	205 (0.2)	4 (0.1)	11 (0.2)	0 (0)	1 (0.1)
*2A/*2A		1 (0)	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)
*1/ rs67376798		787 (0.8)	720 (0.9)	12 (0.4)	35 (0.5)	11 (0.1)	9 (1.2)
rs67376798/ rs67376798		14 (0)	12 (0)	0 (0)	2 (0)	0 (0)	0 (0)
*2A/ rs67376798		2 (0)	2 (0)	0 (0)	0 (0)	0 (0)	0 (0)

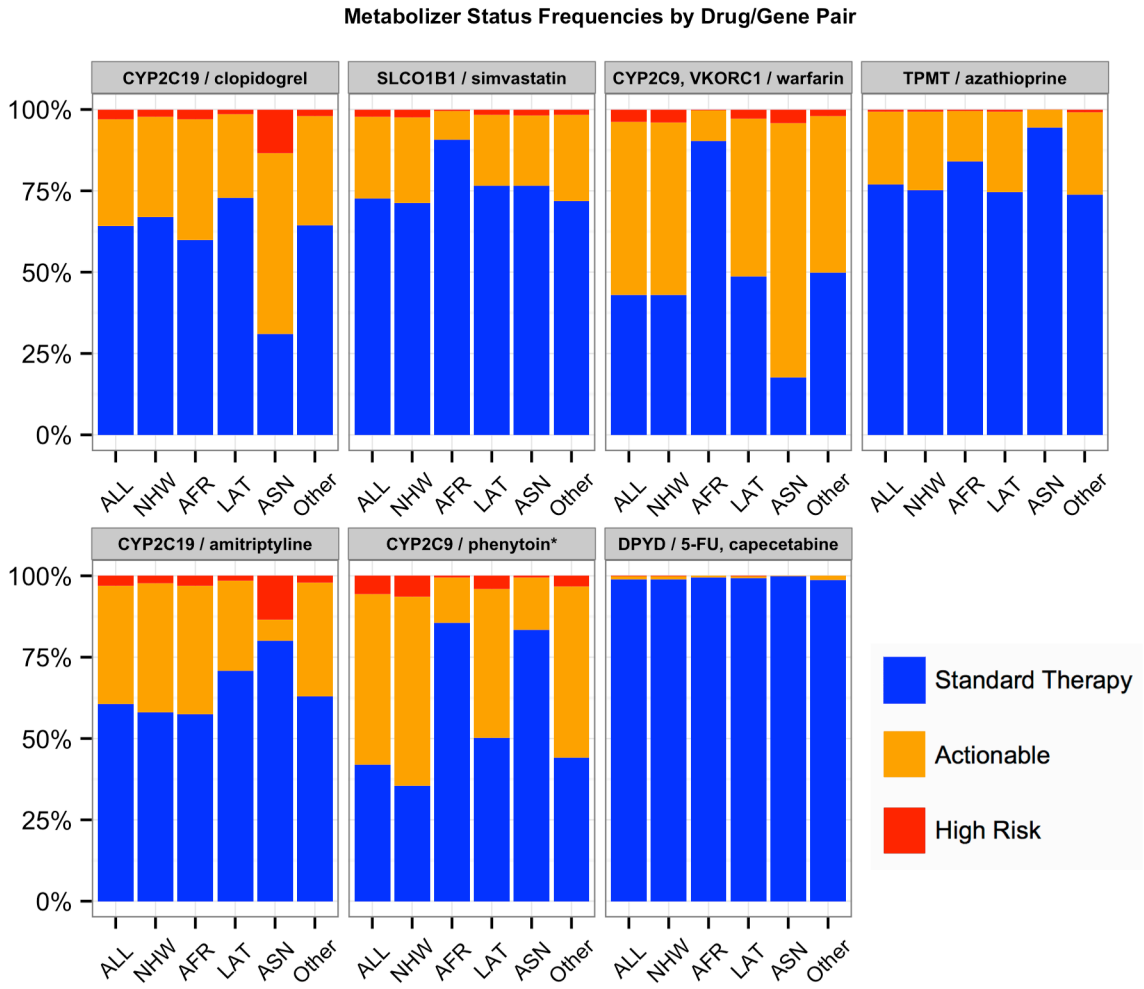
Frequency, by ethnicity, of relevant star allele diplotypes for the 7 drug-gene interactions in this study. Only star alleles with CPIC guidelines are included, as these are used to establish metabolizer phenotypes. NHW = Non-Hispanic White, AFR = African American, LAT = Latino, ASN = Asian.

2.2.3 Metabolizer Status and Clinical Relevance

Based on star allele diplotypes and CPIC guidelines, we assigned each cohort member a metabolizer status for the 7 DGIs. We considered a metabolizer status to be actionable for a DGI if it warranted a change in dose or therapy based on guidelines. For example, individuals with one or more of the CYP2C19 *2 or *3 alleles were considered intermediate or poor metabolizers of clopidogrel, with guidelines suggesting alternate therapy for the 35.7% (N=36,784) of the GERA cohort with these phenotypes. Poor metabolizers, typically with more than one nonfunctional or reduced functional allele, constitute a subset of those with actionable metabolizer status. For most drugs, actionable allele frequencies were highly variable among ethnicities, as shown in Figure 2.2. For 6 of the 7 DGIs, one-versus-all chi-squared tests show that counts of metabolizer status phenotypes vary significantly by ethnicity (adjusted $P < 0.01$ for all groups). For 5-fluorouracil and capecetabine, African American ($\chi^2(2, N=8081) = 9.8$, adjusted $P = 0.04$) and Latino ($\chi^2(2, N=7246) = 10.1$, adjusted $P = 0.03$) metabolizer status counts did not vary significantly from other ethnicities.

We found that overall, 89% of GERA patients (90% of White, 76% African American, 81% Latino, 93% Asian, 88% Other/Uncertain) had at least one actionable allele across these 7 DGIs, while 13% of GERA were categorized as high-risk poor responders (14% of White, 9.2% African American, 8.1% Latino, 16% Asian, and 11% Other/Uncertain). 53% of GERA cohort members had actionable alleles in 2 or more pharmacogenes.

Figure 2.2 Metabolizer status frequency by ethnicity



For each drug-gene interaction, plot shows frequency of cohort members, by ethnicity, who should receive standard therapy, or have actionable or high-risk phentoypes that warrant a change in dose or switch to an alternate medication. ALL = All GERA cohort members, NHW = Non-Hispanic White, AFR = African American, LAT = Latino, ASN = Asian.

To assess the clinical relevance of pharmacogenetic information, we analyzed outpatient pharmacy records for the 7 study drugs by ethnicity (Figure 2.1). Additionally, pharmacy records for 27 other drugs with moderate to strong CPIC evidence for pharmacogenetic-based prescribing were assessed (Table 2.4). 66% of the cohort had at least one valid prescription for an associated drug (68% of White individuals, 66% of African Americans, 62% of Latinos, 51% of Asians, and 72% of Others). 57% of those exposed to a CPIC drug had valid prescriptions for

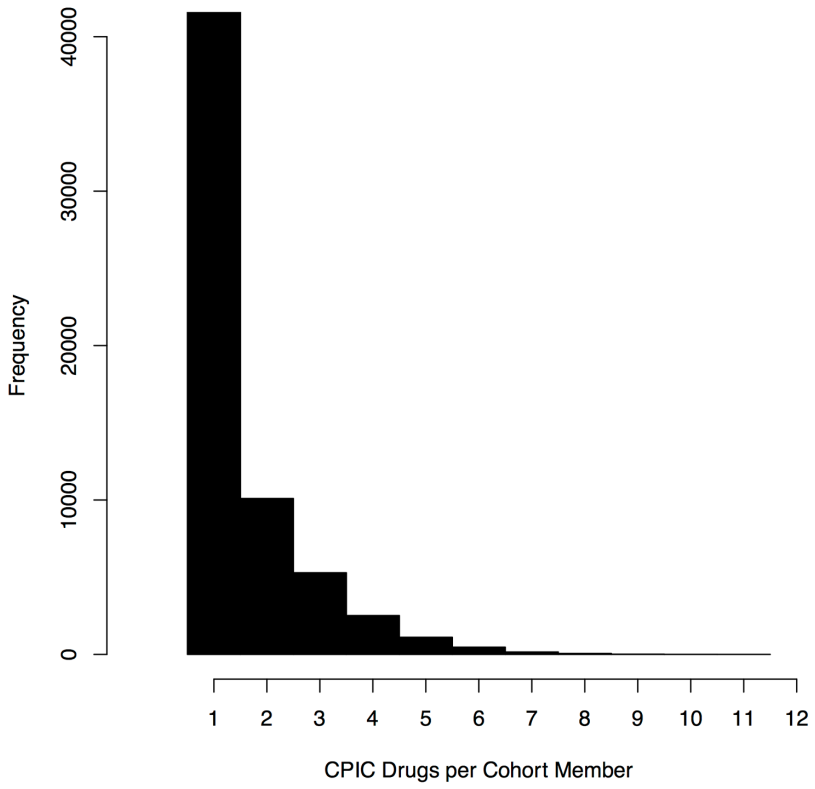
two or more associated drugs (representing 37% of GERA), highlighting the potential impact of pharmacogenetic information in older subjects with a higher likelihood of polypharmacy (Figure 2.3).

Table 2.4 Drugs with level A CPIC guidelines for medication exposure review

Gene	Drug	Pharmacogenetic Information on FDA Label
HLA-B	abacavir	Genetic testing recommended
HLA-B	allopurinol	
CYP2C19	amitriptyline	
CYP2D6	amitriptyline	Actionable PGx
TPMT	azathioprine	Genetic testing recommended
IFNL3	boceprevir	Informative PGx
DPYD	capecitabine	Actionable PGx
HLA-B	carbamazepine	Genetic testing required
CYP2C19	citalopram	Actionable PGx
CYP2C19	clopidogrel	Genetic testing recommended
CYP2D6	codeine	Actionable PGx
CYP2D6	desipramine	Actionable PGx
CYP2D6	doxepin	Actionable PGx
CYP2C19	doxepin	
CYP2C19	escitalopram	
DPYD	fluorouracil	Actionable PGx
CYP2D6	fluvoxamine	Informative PGx
CYP2C19	imipramine	
CYP2D6	imipramine	Actionable PGx
TPMT	mercaptopurine	Genetic testing recommended
CYP2D6	nortriptyline	Actionable PGx
CYP2D6	oxycodone	
CYP2D6	paroxetine	Informative PGx
IFNL3	peginterferon alfa-2b	Actionable PGx
IFNL3	peginterferon alfa-2a	
HLA-B	phenytoin	Actionable PGx
CYP2C9	phenytoin	

G6PD	rasburicase	Genetic testing required
IFNL3	ribavirin	
SLCO1B1	simvastatin	
CYP3A5	tacrolimus	
CYP3A5	tacrolimus	
CYP2D6	tamoxifen	
IFNL3	telaprevir	Actionable PGx
TPMT	thioguanine	Actionable PGx
CYP2D6	tramadol	Actionable PGx
CYP2C19	trimipramine	
CYP2D6	trimipramine	Actionable PGx
CYP2C19	voriconazole	Informative PGx
CYP2C9	warfarin	Actionable PGx
VKORC1	warfarin	Actionable PGx

Figure 2.3 Histogram of number of CPIC drugs prescribed per cohort member



Count of CPIC Level A Drugs prescribed per cohort member. Count based on observation of at least one outpatient prescription for a CPIC drug listed in Table 2.4.

We found that 47% of the GERA cohort had been exposed to at least one of the seven drugs in this study. In total, 17,747 individuals (17% of GERA) had been prescribed at least one study drug for which they had an actionable metabolizer phenotype based on pharmacogenetics. Table 2.5 shows the number of subjects who had prescriptions for drugs in the 7 DGIs as well as actionable or high-risk metabolizer status phenotypes. This represents 18.3% of White individuals, 10% of African Americans, 13% of Latinos, 12% of Asians, and 18% of Others (Table 2.6). Table 2.7 highlights the subset of the of the cohort (4.3%) that had received drugs for which they were poor metabolizers. Overall, 17,747 individuals (17% of GERA) may have been given an incorrect dose or medication, whose treatment may have benefitted from availability of pharmacogenetic information.

Table 2.5 Medication exposure of cohort members with actionable or high-risk phenotypes

Gene	Drug	Total Patients Who Received Drug	% Received Drug with Actionable Phenotype	% Received Drug with High-Risk Phenotype
SLCO1B1	simvastatin	31,096	28%	2.3%
CYP2C9/VKORC1	warfarin	12,863	32%	4.0%
CYP2C19	clopidogrel	6,647	59%	7.4%
DPYD	5-fluorouracil or capecitabine	9,153	1.1%	0.011%
CYP2C19	amitriptyline	7,323	41%*	38%
TPMT	azathioprine	627	8.3%	0.32%
CYP2C9	phenytoin**	566	29%	2.8%

Counts of cohort members with at least one outpatient prescription for study drugs. For each drug-gene interaction, table lists total number of cohort members who were prescribed the drug and have actionable or high-risk phenotypes based on their star allele status for the given gene. Actionable phenotype frequencies are inclusive of high-risk phenotype frequencies. *For amitriptyline, both PMs and UMs are considered actionable phenotypes requiring a change in dose or alternate therapy. Amitriptyline is not considered efficacious in ultra-rapid metabolizers, who were classified as high-risk. PMs were considered to have actionable phenotypes for amitriptyline. **For phenytoin, reported phenotype frequencies are independent of HLA-B*15:02 status, which also plays a role in phenytoin response.

Table 2.6 Actionable phenotypes among cohort members with medication exposure, by ethnicity

			Ethnicity (% of Total Exposed by Ethnicity)				
Drug	Total Exposed	Total Exposed with Actionable Phenotype (% of Exposed)	White	African American	Latino	Asian	Other/ Uncertain
clopidogrel	6647	3944 (59%)	3373 (60%)	128 (64%)	157 (45%)	245 (67%)	41 (63%)
simvastatin	31096	8552 (28%)	7418 (29%)	94 (8%)	491 (23%)	481 (22%)	68 (28%)
warfarin	12863	4148 (32%)	3637 (32%)	19 (5%)	161 (32%)	300 (68%)	31 (26%)
azathioprine	627	52 (8.3%)	46 (9%)	2 (9%)	2 (5%)	2 (5%)	0
amitriptyline	7323	2993 (41%)	2581 (42%)	137 (47%)	180 (32%)	75 (24%)	20 (27%)
capecitabine	440	6 (1.4%)	6 (2%)	0	0	0	0
5-fluorouracil	8713	93 (1.1%)	92 (1%)	0	1 (1%)	0	0
phenytoin	566	162 (29%)	152 (32%)	3 (10%)	6 (16%)	1 (5%)	0

Table 2.7 High-risk phenotypes among cohort members with medication exposure, by ethnicity

			Ethnicity (% of Total Exposed by Ethnicity)				
Drug	Total Exposed	Total Exposed with High-Risk Phenotype (% of Exposed)	White	African American	Latino	Asian	Other/ Uncertain
clopidogrel	6647	496 (7.4%)	413 (7%)	12 (6%)	14 (4%)	55 (15%)	2 (3%)
simvastatin	31096	717 (2.3%)	633 (2%)	5 (0.44%)	35 (2%)	41 (2%)	3 (1%)
warfarin	12863	518 (4.0%)	487 (4%)	0	14 (3%)	14 (3%)	3 (3%)
azathioprine	627	2 (0.32%)	2	0	0	0	0

			(0.4%)				
amitriptyline	7323	2798 (38%)	2448 (40%)	130 (44%)	172 (30%)	29 (9%)	19 (25%)
capecitabine	440	0	0	0	0	0	0
5-fluorouracil	8713	1 (0.011%)	1 (0.01%)	0	0	0	0
phenytoin	566	16 (2.8%)	14 (3%)	0	2 (5%)	0	0

2.3 Materials & Methods

2.3.1 Study Population

Participants in this study are members of the Kaiser Permanente RPGEH GERA cohort (N=102,979). A detailed description of the GERA cohort has been previously published[7] and can also be found in dbGAP phs000674.v1.p1. At the time of enrollment, GERA members were adult members of Kaiser Permanente Medical Care Plan, Northern California Region with high-density genotype data linked to Kaiser Permanente electronic health data. EHR data included diagnoses, laboratory, procedure, and pharmacy utilization records for all members, as well as survey data on lifestyle and environmental factors based on residence information. Prescription data availability for the GERA cohort began in January 1, 1995. Pharmacy records were queried up to and including December 31, 2014. Thus the maximum medication exposure time with data available for an individual in the cohort who has continuously been a KPNC member since prior to 1995 would be 20 years. RPGEH enrollment was voluntary and all members broadly consented to use of their data for research on health and disease. The study was reviewed and approved by the UCSF Committee on Human Research, as well as the Kaiser Permanente Institutional Review Board.

2.3.2 Genotyping and Ancestry

Genotyping was performed using four custom Affymetrix Axiom arrays created for the four major race-ethnicity groups in the cohort: African Americans, East Asians, Latinos, and

Non-Hispanic Whites[18,19]. Arrays included 674-893k SNPs, depending on the array[18,19]. Arrays were designed to maximize genome-wide coverage of low frequency and common variants in each group. Self-reported race/ethnicity was generally used to assign subjects to arrays, and used to define ethnicities in this study. Genotyping was performed by the Genomics Core Facility of the Institute for Human Genetics at UCSF using standard and enhanced Affymetrix Gene Titan Axiom protocols[7]. Further details about race/ethnicity assignment, principal components analysis, genotyping, and quality control procedures have been previously published[7,20].

2.3.3 Pharmacogene Selection and Actionability

An initial list of relevant pharmacogenes was compiled from the PharmaADME.org Core List[21], PharmGKB Very Important Genes[22], and the FDA Table of Pharmacogenomic Biomarkers in Drug Labels[23]. The list included drug-metabolizing enzymes, transporters, and some target genes and receptors. Drugs associated with genes with Level A CPIC guidelines[24,25] and outpatient pharmacy prescriptions for at least 500 GERA cohort members were selected for further review in this analysis. For the 7 selected DGIs, high quality, phased genotype data was also available for all star alleles with CPIC guidelines within the gene of interest.

Actionability was determined by the presence of moderate to strong clinical evidence that supports the use of genetic information to change prescribing of an affected drug, as determined by CPIC[9]. In this study, we consider a pharmacogenetic phenotype to be actionable if it warrants a change in prescribing: either a change in dose or an alternate therapy.

2.3.4 Genotype Imputation

After genotyping quality control, SNPs with overall call rates <90% were removed prior to imputation. Imputation was performed separately for each array by pre-phasing the genotypes with SHAPEIT v2.r727[26], including cryptic-related first-degree relatives modeled to aid in

phasing. Variants were then imputed from the 1000 Genomes Project[27] as a cosmopolitan reference panel with IMPUTE2 v2.3.0[28,29]. The estimated quality control r^2 metric from Impute2 showed a high estimate of correlation of the imputed genotype to the true genotype[30] for all SNPs with CPIC guidelines in this analysis ($r^2 > 0.8$ for majority of SNPs, shown in Table 2.8). For pharmacogenes of interest, phased, imputed genotype probabilities for all SNPs associated with documented star alleles were extracted for use in this analysis.

Table 2.8 Imputation r^2 for study SNPs by array

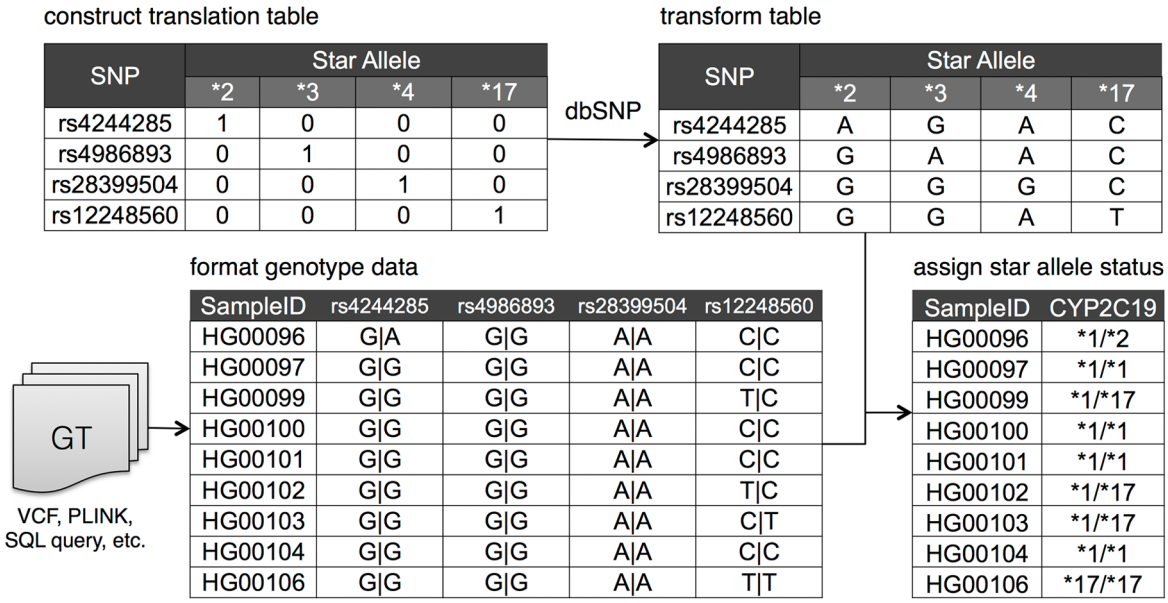
Gene	SNP	Star Alelle(s)	EUR	AFR	LAT	ASN
CYP2C19	rs4244285	*2	1.00	1.00	1.00	1.00
CYP2C19	rs4986893	*3	0.98	1.00	0.99	1.00
CYP2C19	rs12248560	*17	0.97	0.99	0.98	0.99
CYP2C9	rs1799853	*2	1.00	1.00	1.00	0.98
CYP2C9	rs1057910	*3	1.00	1.00	1.00	1.00
DPYD	rs3918290	*2A	0.67	0.67	0.64	0.46
DPYD	rs67376798	rs67376798	0.85	0.80	0.88	0.52
SLCO1B1	rs4149056	*5, *15, *17	1.00	1.00	1.00	1.00
TPMT	rs1800462	*2	0.97	1.00	0.56	1.00
TPMT	rs1800460	*3A, *3B	0.97	0.89	0.94	0.76
TPMT	rs1142345	*3A, *3C	1.00	1.00	1.00	0.99
VKORC1	rs9923231	*2	1.00	1.00	1.00	0.99

2.3.5 Translation Table Lookup

Multiple pharmacogenetic resources were consulted in order to construct translation tables for each gene, including the Human Cytochrome P450 Allele Nomenclature Database[31], PharmGKB[32], and individual CPIC guidelines for each DGI. Binary translation tables were constructed to indicate the presence or absence of specific SNP alleles in a relevant star haplotype (Table 2.9). NCBI's dbSNP[33] was used to create allele-based tables, which were

consulted in the process of assigning star diplotypes to each cohort member based on their phased genotype data for the relevant SNPs (Figure 2.4).

Figure 2.4 Workflow diagram.



SNP	Star Allele			
	*2	*3	*4	*17
rs4244285	1	0	0	0
rs4986893	0	1	0	0
rs28399504	0	0	1	0
rs12248560	0	0	0	1

SNP	Star Allele			
	*2	*3	*4	*17
rs4244285	A	G	A	C
rs4986893	G	A	A	C
rs28399504	G	G	G	C
rs12248560	G	G	A	T

SampleID	rs4244285	rs4986893	rs28399504	rs12248560
HG00096	G|A	G|G	A|A	C|C
HG00097	G|G	G|G	A|A	C|C
HG00099	G|G	G|G	A|A	T|C
HG00100	G|G	G|G	A|A	C|C
HG00101	G|G	G|G	A|A	C|C
HG00102	G|G	G|G	A|A	T|C
HG00103	G|G	G|G	A|A	C|T
HG00104	G|G	G|G	A|A	C|C
HG00106	G|G	G|G	A|A	T|T

SampleID	CYP2C19
HG00096	*1/*2
HG00097	*1/*1
HG00099	*1/*17
HG00100	*1/*1
HG00101	*1/*1
HG00102	*1/*17
HG00103	*1/*17
HG00104	*1/*1
HG00106	*17/*17

Translation tables were compiled for 7 DGIs from various pharmacogenetic resources and dbSNP (see Materials & Methods). Tables were then used to convert phased genotype data into star allele diplotypes, which can be used for metabolizer status phenotype assignment.

Translation tables were created to aid conversion of dbSNP rs identifiers to haplotype-based star alleles. Tables were compiled using the Human Cytochrome P450 Allele Nomenclature Database[5], PharmGKB[32], and specific CPIC guidelines to indicate the presence or absence of an rs identified SNP allele in the star allele haplotype (Table 2.9). DbSNP build 141 was used to identify specific reference and alternate alleles. The compiled translation tables were referenced to convert phased genotype data into star allele diplotypes for each sample. All analyses were performed using R version 3.0.2[34].

Table 2.9 Translation tables for star alleles in the 7 study DGIs

SNP	CYP2C19 *2	CYP2C19 *3	CYP2C19 *17	CYP2C9 *2	CYP2C9 *3	DYPD *2A	DPYD rs67376798	SLCO1B1 *5	TPMT *2	TPMT *3A	TPMT *3B	TPMT *3C	VKORC1 *2
rs4244285	1	0	0	0	0	0	0	0	0	0	0	0	0
rs4986893	0	1	0	0	0	0	0	0	0	0	0	0	0
rs12248560	0	0	1	0	0	0	0	0	0	0	0	0	0
rs1799853	0	0	0	1	0	0	0	0	0	0	0	0	0
rs1057910	0	0	0	0	1	0	0	0	0	0	0	0	0
rs3918290	0	0	0	0	0	1	0	0	0	0	0	0	0
rs67376798	0	0	0	0	0	0	1	0	0	0	0	0	0
rs4149056	0	0	0	0	0	0	0	1	0	0	0	0	0
rs1800462	0	0	0	0	0	0	0	0	1	0	0	0	0
rs1800460	0	0	0	0	0	0	0	0	0	1	1	0	0
rs1142345	0	0	0	0	0	0	0	0	0	1	0	1	0
rs9923231	0	0	0	0	0	0	0	0	0	0	0	0	1

Note that while many actionable star alleles represent reduced gene function, not all star alleles result in the same metabolizer phenotype for all associated drugs. In our case, CPIC guidelines suggest that individuals heterozygous for CYP2C19*2 or *3 (i.e. one copy) are poor metabolizers of clopidogrel. If they should need antiplatelet therapy, CPIC guidelines recommend alternative drugs, such as prasugrel and ticagrelor. While these cohort members are considered poor metabolizers for clopidogrel, they are also considered to be extensive metabolizers of amitriptyline, which warrants standard therapy for this drug. While CYP2C19 is a known metabolizer of amitryptyline, only those with 2 copies of these reduced function alleles (*2 or *3) are considered to be poor metabolizers of amitryptyline.

Based on star allele diplotypes and CPIC guidelines, we assigned each cohort member a metabolizer status for the 7 DGIs. We considered a metabolizer status to be actionable for a DGI if it warranted a change in dose or therapy based on guidelines. For example, individuals with one or more of the CYP2C19 *2 or *3 alleles were considered intermediate or poor metabolizers of clopidogrel, with guidelines suggesting alternate therapy for the GERA cohort members with these phenotypes. Poor metabolizers, typically with more than one nonfunctional or reduced function allele, constitute a (more severe) subset of those with actionable metabolizer status.

2.3.6 Pharmacy Data

Complete pharmacy records began in 1995 and include coded entries for all outpatient and inpatient prescriptions. For the selected pharmacogenes, 7 associated drugs were selected for pharmacy record review, based on frequency among GERA cohort members and CPIC Level A evidence for the drug-gene pair, suggesting that genetic information should be used to change prescribing of the affected drug[9,25]: CYP2C19 (clopidogrel, amitriptyline, citalopram), CYP2C9 (phenytoin) and VKORC1 (warfarin), SLCO1B1 (simvastatin), TPMT (azathioprine), and

DPYD (capecitabine, 5-fluorouracil). Outpatient pharmacy records in the KPNC research database were queried. For each drug, the generic label name included the drug of interest as an active ingredient. Pharmacy records were filtered to include only those picked up (not returned-to-stock) by GERA cohort members.

2.4 Discussion

The GERA cohort represents one of the largest genetic datasets linked to a real-world, observational EHR resource with two decades of pharmacy information in a single healthcare network. While genetic data in our study were not generated under Clinical Laboratory Improvement Amendments (CLIA)-approved protocols, and thus cannot be clinically used to change prescribing for patients, they allowed us to perform a preliminary assessment of the utility of pharmacogenetic testing in an actual healthcare setting. If DNA were broadly available, it would be possible to preemptively genotype patients for known drug-gene pairs before these drugs are administered[2]. Alternatively, patients in race-ethnicity groups known to have increased frequency of genetic variants conferring risk of adverse drug responses, as well as those in age or diagnostic groups likely to be prescribed drugs known to interact with these genes, could be preemptively genotyped using a pharmacogenetic test panel spanning multiple genetic variants. Theoretically, for those most likely to be impacted by pharmacogenetic information, data would be available in the EHR at the time a prescription is written by a physician. Thus, a major motivation of this study was to determine the frequency of genetic variation and interactions with medications that carry potential clinical consequences in terms of drug response and adverse events in a large, diverse population in a clinical context. This understanding is vital to translating findings into clinically relevant actions, such as making pharmacogenetic testing available, and changing prescribing patterns and dosing to optimize drug therapy. For many

well-studied drugs, this genetic information can be utilized to predict how patients will respond before they are prescribed a drug[35]. Physicians can then consider alternate therapies or dosing for non-responders, which may avoid adverse outcomes[36].

In this study, we found that 89% of the 102,979 adult cohort members in GERA had pharmacogenetic alleles that could influence a treatment decision for 7 different drugs. While this number is relatively large, those with actionable or high-risk phenotypes that were actually given at least one of the relevant drugs, while they were members of KPNC, was approximately 17% of the cohort. Given the median age of the cohort (64 years), these estimates are likely an underestimate of lifetime medication exposure in this population, since many younger individuals in the cohort are less likely to have prescriptions for several of the drugs in this study but may in the future. Based on the observed data, this 17% represents the minimum impact that preemptive pharmacogenetic genotyping may have in this population for leading to actionable therapy changes. Further investigation is necessary to fully address the benefits of pharmacogenetic information, especially in different age groups. Additionally, quantifying the cost of any adverse drug events or hospitalizations in this subset of 17,747 people may show that preemptive genotyping for pharmacogenetic application is worth the cost of implementation. Collecting additional data on diagnosis prevalence for the diseases these drugs are used to treat, as well as information about alternative treatments, are also important factors in assessing cost-benefit of preemptive genotyping.

Our study also highlights the potential of this resource for pharmacogenetic discovery, particularly in ethnically diverse populations. Many traditional pharmacogenetic studies were based on randomized clinical trials (RCTs) studying treatment effect in rigorously defined populations[37]. While such studies employ high quality data collection and quality control

measures, the studies can be costly both in time and money, may have limited follow-up periods, and may exclude several patient populations. For instance, many pharmacogenetic studies have not included minority populations in their analyses. The GERA cohort, which includes thousands of non-Caucasian individuals, allowed us to perform a preliminary assessment of variability among multiple ethnic groups in pharmacogenetic variants of known significance. Following up on this assessment with phenotype data derived from the EHR may help us to better understand how genetics influence drug response in understudied populations. Additionally, since impact of ethnicity may not be known for many drugs until the postmarketing stage[38], this resource may be of particular use in pharmacovigilance studies. Since the data has already been collected, the EHR also represents a reusable resource for extracting various phenotypes, with potentially longer follow-up periods, than can be accomplished in RCTs.

Despite the potential research and clinical impact of such a resource, numerous additional hurdles prevent widespread implementation of preemptive pharmacogenetic genotyping. At the sample level, collection of genetic information or extraction from biobanked samples must be performed under CLIA-approved protocols in order to be utilized in clinic. While the RPGEH genetic data was generated for research purposes, the 7 drug-gene pairs assessed in this study represent genes for which commercially-available pharmacogenetic tests currently exist. Once genetic data is generated, conversion to clinically-relevant star allele nomenclature must be performed. While our study utilized currently published translation tables and literature to perform this task, such resources should not be considered static. With the adoption of next-generation sequencing technologies, future studies may reveal less common functional variants of clinical significance to drug response[38]. Translation resources and knowledge bases must be maintained to reflect current clinical evidence. From a technology standpoint, presenting

pharmacogenetic information at point-of-care is a non-trivial task that incorporates clinical decision support (CDS) updates, as well as physician education regarding the impact of genetic information on prescribing decisions[39-41]. Some institutions have successfully implemented pharmacogenetic CDS for certain drug-gene pairs and have highlighted detailed technical and institutional challenges associated with this task[3,35,42,43]. At an institutional level, cost-benefit analyses must be performed to fully quantify the costs associated with preemptive pharmacogenetic genotyping[36,44], including sample collection, genotyping and data storage, reimbursement, infrastructure changes or development, maintenance of translational resources and prescribing guidelines, as well as prescriber training and education[39,41,45].

Independent of implementation hurdles, the real-world application of such a resource highlights further challenges. For a large health care system such as KPNC, investing in preemptive pharmacogenetic genotyping may be more impactful in the long-term compared to smaller, more segregated health networks. For our study, our cohort had a median of 14.9 years of pharmacy records available for drugs with CPIC guidelines, suggesting that early genotyping of patients may be worth the upfront costs if most members stay in-network long enough for pharmacogenetic information to be utilized. For other health networks, questions of data ownership may arise, as information derived in one system may not be portable to another system. As patient-centered initiatives grow, patients may wish to own their data, especially pharmacogenetic information that may be relevant to them at a future date.

2.5 References

1. Relling, M. V. & Klein, T. E. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther* **89,** 464–467 (2011).

2. Bielinski, S. J. *et al.* Preemptive Genotyping For Personalized Medicine: Design Of The Right Drug, Right Dose, Right Time-Using Genomic Data To Individualize Treatment Protocol. *Mayo Clin Proc* **89,** 25–33 (2014).

3. Pulley, J. M. J. *et al.* Operational Implementation Of Prospective Genotyping For Personalized Medicine: The Design Of The Vanderbilt PREDICT Project. *Clin Pharmacol Ther* **92,** 87–95 (2012).

4. Robarge, J. D., Li, L., Desta, Z., Nguyen, A. & Flockhart, D. A. The Star-Allele Nomenclature: Retooling For Translational Genomics. *Clin Pharmacol Ther* **82,** 244–248 (2007).

5. Sim, S. C. & Ingelman-Sundberg, M. The Human Cytochrome P450 (CYP) Allele Nomenclature Website: A Peer-Reviewed Database Of CYP Variants And Their Associated Effects. *Hum Genomics* **4,** 278–281 (2010).

6. Potamias, G. *et al.* Deciphering Next-Generation Pharmacogenomics: An Information Technology Perspective. *Open Biology* **4,** 140071 (2014).

7. Kvale, M. N. *et al.* Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200,** 1051–1060 (2015).

8. Hajjar, E. R., Cafiero, A. C. & Hanlon, J. T. Polypharmacy In Elderly Patients. *Am J Geriatric Pharmacother* **5,** 345–351 (2007).

9. Caudle, K. E. *et al.* Incorporation Of Pharmacogenomics Into Routine Clinical Practice: The Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline Development Process. *Curr Drug Metab* **15,** 209–217 (2014).

10. Xie, H.-G., Prasad, H. C., Kim, R. B. & Stein, C. M. CYP2C9 Allelic Variants: Ethnic

Distribution And Functional Significance. *Adv Drug Deliv Rev* **54,** 1257–1270 (2002).

11. Scott, S. A. *et al.* Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C19 Genotype and Clopidogrel Therapy: 2013 Update. *Clin Pharmacol Ther* **94,** 317–323 (2013).

12. Wilke, R. A. *et al.* The Clinical Pharmacogenomics Implementation Consortium: CPIC Guideline For SLCO1B1 And Simvastatin-Induced Myopathy. *Clini Pharmacol Ther* **92,** 112–117 (2012).

13. Relling, M. V. *et al.* Clinical Pharmacogenetics Implementation Consortium Guidelines For Thiopurine Methyltransferase Genotype And Thiopurine Dosing. *Clin Pharmacol Ther* **89,** 387–391 (2011).

14. Johnson, J. A. *et al.* Clinical Pharmacogenetics Implementation Consortium Guidelines For CYP2C9 And VKORC1 Genotypes And Warfarin Dosing. *Clin Pharmacol Ther* **90,** 625–629 (2011).

15. Hicks, J. K. *et al.* Clinical Pharmacogenetics Implementation Consortium Guideline For CYP2D6 And CYP2C19 Genotypes And Dosing Of Tricyclic Antidepressants. *Clin Pharmacol Ther* **93,** 402–408 (2013).

16. Caudle, K. E. *et al.* Clinical Pharmacogenetics Implementation Consortium Guidelines For Dihydropyrimidine Dehydrogenase Genotype And Fluoropyrimidine Dosing. *Clin Pharmacol Ther* **94,** 640–645 (2013).

17. Caudle, K. E. *et al.* Clinical Pharmacogenetics Implementation Consortium Guidelines For CYP2C9 And HLA-B Genotypes And Phenytoin Dosing. *Clin Pharmacol Ther* **96,** 542–548 (2014).

18. Hoffmann, T. J. *et al.* Design And Coverage Of High Throughput Genotyping Arrays

Optimized For Individuals Of East Asian, African American, And Latino Race/Ethnicity Using Imputation And A Novel Hybrid SNP Selection Algorithm. *Genomics* **98,** 422–430 (2011).

19. Hoffmann, T. J. *et al.* Next Generation Genome-Wide Association Tool: Design And Coverage Of A High-Throughput European-Optimized SNP Array. *Genomics* **98,** 79–89 (2011).

20. Banda, Y. *et al.* Characterizing Race/Ethnicity And Genetic Ancestry For 100,000 Subjects In The Genetic Epidemiology Research On Adult Health And Aging (GERA) Cohort. *Genetics* **200,** 1285–1295 (2015).

21. PharmaADME.org. Core Marker List. http://pharmaadme.org/joomla/index.php?option=com_content&task=view&id=14&Itemid=29 (2015).

22. PharmGKB. VIPs: Very Important Pharmacogenes. https://www.pharmgkb.org/view/vips.jsp?browseKey=vipGenes (2015).

23. U.S. Food and Drug Administration. Table of Pharmacogenomic Biomarkers in Drug Labels. http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm (2015).

24. PharmGKB. CPIC Genes/Drugs. https://www.pharmgkb.org/cpic/pairs (2015).

25. PharmGKB. CPIC Genes/Drugs Legend. https://www.pharmgkb.org/page/cpicStatusLegend (2015).

26. Delaneau, O., Marchini, J. & Zagury, J.-F. A Linear Complexity Phasing Method For Thousands Of Genomes. *Nat Methods* **9,** 179–181 (2012).

27. McVean, G. A. *et al.* An Integrated Map Of Genetic Variation From 1,092 Human Genomes. *Nature* **491,** 56–65 (2012).

28. Howie, B. N., Donnelly, P. & Marchini, J. A Flexible And Accurate Genotype Imputation Method For The Next Generation Of Genome-Wide Association Studies. *PLoS Genet* **5**, e1000529 (2009). doi: 10.1371/journal.pgen.1000529

29. Howie, B. N., Marchini, J. & Stephens, M. Genotype Imputation with Thousands of Genomes. *Genes|Genomes|Genetics* **1,** 457–470 (2011). doi: 10.1534/g3.111.001198

30. Marchini J. & Howie, B. Genotype Imputation for Genome-Wide Association Studies. *Nat Rev Genet* **11**, 499-511 (2010).

31. Sim, S. C. Human Cytochrome P450 (CYP) Allele Nomenclature Database. http://www.cypalleles.ki.se/ (2014).

32. Whirl-Carrillo, M. *et al.* Pharmacogenomics Knowledge for Personalized Medicine. *Clin Pharmacol Ther* **92,** 414–417 (2012).

33. Sherry, S. T. *et al*. Dbsnp: The NCBI Database Of Genetic Variation. *Nucleic Acids Res* **29,** 308-311 (2001).

34. R Core Team. R: A Language And Environment For Statistical Computing. R Foundation For Statistical Computing. http://www.r-project.org/ (2015).

35. Bell, G. C. *et al.* Development And Use Of Active Clinical Decision Support For Preemptive Pharmacogenomics. *J Am Med Inform Assoc* (2013). doi:10.1136/amiajnl-2013-001993

36. Wu, A. C. & Fuhlbrigge, A. L. Economic Evaluation Of Pharmacogenetic Tests. *Clin Pharmacol Ther* **84,** 272–274 (2008).

37. Bowton, E. *et al.* Biobanks And Electronic Medical Records: Enabling Cost-Effective

Research. *Sci Transl Med* **6,** 234cm3 (2014).

38. Yasuda, S. U., Zhang, L. & Huang, S.-M. The Role Of Ethnicity In Variability In Response To Drugs: Focus On Clinical Pharmacology Studies. *Clin Pharmacol Ther* **84,** 417–423 (2008).

39. Crews, K. R., Hicks, J. K., Pui, C. H., Relling, M. V. & Evans, W. E. Pharmacogenomics And Individualized Medicine: Translating Science Into Practice. *Clin Pharmacol Ther* (2012). doi:10.1038/clpt.2012.120

40. Scott, S. A. Personalizing Medicine With Clinical Pharmacogenetics. *Genet Med* **13,** 987–995 (2011).

41. Rasmussen-Torvik, L. J. *et al.* Design and Anticipated Outcomes of the eMERGE-PGx Project: A Multicenter Pilot for Preemptive Pharmacogenomics in Electronic Health Record Systems. *Clin Pharmacol Ther* **96,** 482–489 (2014).

42. Van Driest, S. L. *et al.* Clinically Actionable Genotypes Among 10,000 Patients With Preemptive Pharmacogenomic Testing. *Clin Pharmacol Ther* **95,** 423–431 (2014).

43. Weitzel, K. W. *et al.* Clinical Pharmacogenetics Implementation: Approaches, Successes, And Challenges. *Am J Med Genet* **166,** 56–67 (2014).

44. Chalkidou, K. & Rawlins, S. M. Pharmacogenetics And Cost-Effectiveness Analysis: A Two-Way Street. *Drug Discov Today* **16,** 873–877 (2011).

45. Baudhuin, L. M., Langman, L. J. & O'Kane, D. J. Translation of Pharmacogenetics into Clinically Relevant Testing Modalities. *Clin Pharmacol Ther* **82,** 373–376 (2007).

3 EHR-DERIVED DRUG RESPONSE: WEIGHT GAIN AND ATYPICAL ANTIPSYCHOTICS

3.1 Introduction

This chapter explores the utility of electronic health records (EHRs) for deriving a drug response phenotype using prescription data and commonly measured weight values. Second-generation antipsychotics, also referred to as atypical antipsychotics (AAPs), are commonly used to treat schizophrenia, bipolar disorder, acute mania, and other psychiatric conditions. The mechanism of action for this class of drug varies, but primarily involves dopamine receptor binding, as well as binding to one or more serotonin receptors. While AAPs are generally associated with fewer extrapyramidal side effects than first-generation antipsychotics, this drug class has been associated with substantial weight gain[1]. AAP-induced weight gain (AIWG), while perhaps less severe than other adverse drug events, is the leading cause for non-compliance among patients who take AAPs[2]. While increased weight has been associated with improved efficacy[3], this side effect may also increase risk for other metabolic or cardiovascular disorders.

The amount of weight gain differs substantially by AAP. For instance, Clozaril (clozapine) and Zyprexa (olanzapine) both pose a higher risk of weight gain than Abilify (aripiprazole) or Geodone (ziprasidone)[2]. Other factors associated with weight gain include lower baseline weight, whether or not the patient has previously taken an AAP, dose and therapy duration[4], and gender[5].

Additional studies have assessed genetic factors underlying AIWG. The majority of such studies were based on candidate genes involved in neurological response or metabolic phenotypes. These include *HTR2C* and *MC4R* from the serotenergic and melanocortin receptor pathways, respectively, as well as the leptin gene *LEP* and obesity-related genes such as *FTO*[6].

Early genetic studies that assessed AAP safety and efficacy were based on data from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study, a blinded study that evaluated five AAPs: olanzapine, quetiapine, risperidone, perphenazine, and ziprasidone[7]. 756 CATIE subjects with genetic information were analyzed in a candidate gene study of treatment response, AIWG, and other response phenotypes[8]. Another study used a candidate gene approach to assess AIWG in other clinical populations[9]. Genome-wide association studies (GWAS) of AIWG have been less common. One GWAS of metabolic phenotypes in CATIE subjects identified SNPs in a small number of metabolism-related genes that were significantly associated, although many more of their top SNPs fell in intergenic regions[10]. In another GWAS of 139 AAP-naïve pediatric patients being treated with aripiprazole, quetiapine, or risperidone, SNPs near *MC4R*, which has previously been associated with BMI and obesity, were found to be associated with AIWG[11]. However, this finding may be AAP-specific, as it was not as strongly replicated in a study of clozapine-treated patients[12].

While previous studies have shown some genetic signal underlying AIWG, they have been limited to candidate gene studies and small population sizes. Large-scale GWAS may provide further insight into other variants associated with this phenotype. Electronic health records (EHRs) provide a means for capturing this side effect in a clinical setting across a larger and broader range of patients, especially since weight is commonly measured during patient visits. In this study, we use EHR data for GERA cohort members as a novel way of quantifying AIWG. We use the extracted phenotype and related covariates to perform a GWAS of BMI changes induced by AAP use.

3.2 Results

3.2.1 AAP Prescriptions

3309 Non-Hispanic White cohort members in GERA had at least one outpatient prescription for an AAP, prior to phenotype filtering by BMI values. Quetiapine was the most commonly prescribed AAP in the GERA cohort (N=1793), followed by risperidone (N=1240), olanzapine (N=1160), aripiprazole (N=800), ziprasidone (N=246), and clozapine (N=39). Less than 10 people were prescribed paliperidone or iloperidone. The average consecutive dose observed was 2-3 months. 65% of patients were observed to be on a single AAP, while 35% had records of 2 or more AAPs.

3.2.2 Subject Demographics

Of the 3309 with outpatient AAP prescriptions, 823 subjects remained after phenotype selection criteria regarding prescriptions and BMI, as described in Materials & Methods, and make up the cohort for this chapter. Cohort demographics and treatment characteristics are described in Table 3.1. All subjects had at least one prescription for the five AAPs assessed in this study, regardless of underlying diagnosis. Overall, the mean age at time of first prescription was 64 years old (s.d. 16 years). All study subjects self-identified as Non-Hispanic White, and were run on the European genotyping array.

3.2.3 Phenotype

For each study subject, prescription records and BMI measurements were ordered according to a medication timeline, as shown in Figure 3.1. The number of BMI measurements was highly variable among subjects, with mean of 13.2 baseline measurements (s.d. 12.7) available prior to AAP therapy, and 2.5 (s.d. 2.3) BMI measurements available during the dosage period. 3 year median baseline BMI did not vary significantly between patients who ultimately

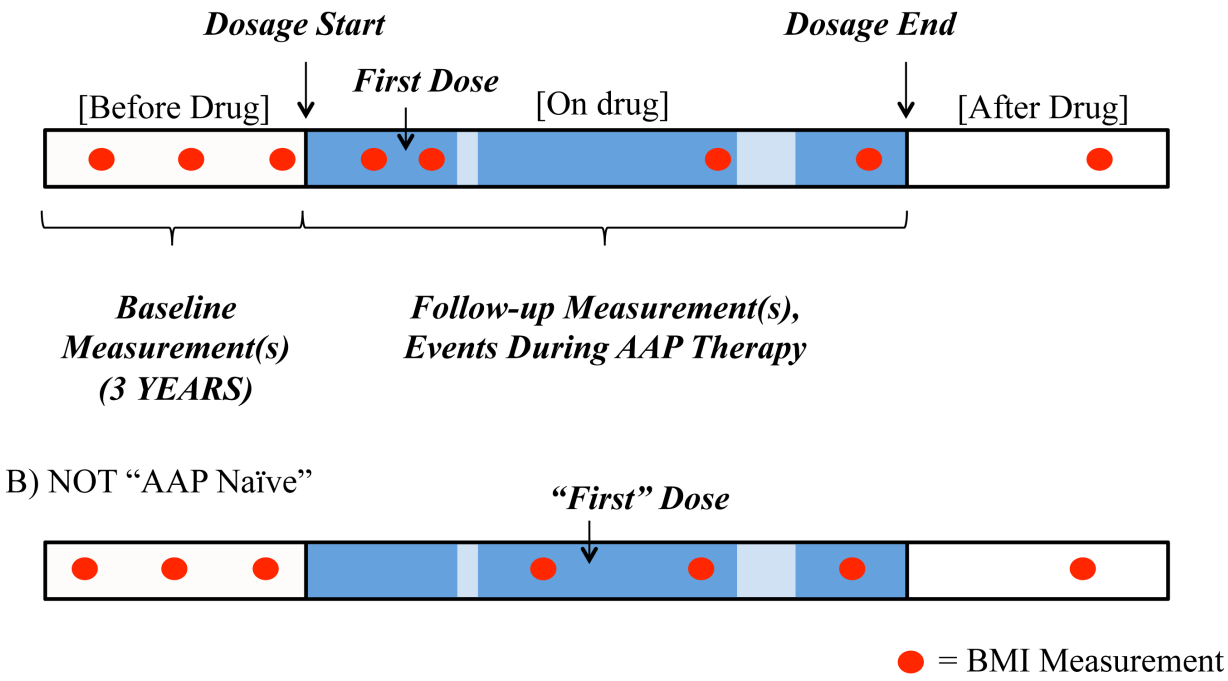
lost weight and gained weight, $t(780) = 0.19$, $P = 0.85$. However, median BMI during AAP therapy was higher in patients who ultimately gained weight compared to those who lost weight, as expected ($t(734) = 7.1$, $P < 0.001$). Patients who gained weight had a median BMI during therapy of 27.8, while patients who lost weight had a median of 25.1.

Table 3.1 Cohort demographics, treatment characteristics, and covariates.

	Mean	SD
Age at Prescription (years)	63.7	15.8
Baseline BMI Measurements	13.2	12.7
BMI Measurements During AAP	2.5	2.3
	N	Percentage
Total	823	--
Male	273	33.2%
AAP Naïve	500	60.8%
AAP: quetiapine	457	55.5%
AAP: aripiprazole	152	18.5%
AAP: risperidone	117	14.2%
AAP: olanzapine	95	11.5%
AAP: clozapine	2	0.243%
Type 2 Diabetic	128	15.5%
Smoker	390	47.4%

Description of 823 Non-Hispanic White cohort members analyzed in this study. BMI measurements refer to the count of BMI data points available per patient during the time period assessed. AAP counts are for the first available AAP with BMI measurements available. That is, if a patient received a different AAP but did not have BMI measurements taken during that dosage period, the patient was labeled not AAP naïve and only the AAP with BMI measurements was recorded.

Figure 3.1 Phenotype timeline schematic.



Baseline BMI for each subject was calculated as the median BMI measurement in the 3 years prior to the AAP dosage start date. A) For AAP naïve patients, BMI measurements were available during their first AAP dosage period. The median of these measurements represents the patient's BMI during AAP therapy. B) Patients who had previously received an AAP but were lacking BMI measurements during that first dosage period were considered not AAP naïve. For these patients, BMI during AAP therapy was calculated as the median BMI value for the first AAP dosage period in which BMI measurements were available. Change in BMI for A) and B) was calculated by subtracting the median baseline BMI from the median BMI during AAP therapy.

Change in BMI was assessed using the two phenotype definitions described in Materials & Methods. Figure 3.2 shows a histogram of the change in BMI (mean -0.12, s.d. 2.5). 382 subjects (46%) showed an increase in BMI, with a mean increase of 1.45 BMI units. 441 (54%) showed a decrease in BMI, with a mean decrease of 1.52 BMI units. Figure 3.3 shows the difference in BMI between the baseline and treatment measurement periods for each subject. Median baseline BMI was highly correlated with median BMI during the dosage period for each subject, $r(821) = 0.93$, $P < 0.001$, as shown in Figure 3.4. 99 subjects showed an increase of over

7% baseline BMI, a cutoff often used to define AIWG in clinical studies. 129 lost over 7% baseline BMI.

Table 3.2 highlights the percentage of people who shifted to a different weight category when they were on AAP treatment. We observed that 15% of those who started with a normal BMI (18.5-24.9 BMI) prior to AAP therapy gained enough weight to be classified as overweight or obese while they were on treatment. Conversely, 17% of those who started with an overweight or obese BMI shifted to a lower BMI classification while they were on treatment.

We observed that females were more likely to gain weight than males (mean 0.03 BMI versus -0.49 BMI), $t(503) = 3.0$, $P < 0.01$. Additionally, patients with Type 2 Diabetes lost more weight than those without the disease (-0.69 BMI versus -0.04 BMI), $t(162) = -1.1$, $P < 0.01$. Smoking status did not have a significant effect on change in BMI, $t\ (816)= 0.35$, $P = 0.73$. Figure 3.5 shows a correlation matrix plot of BMI phenotypes and other numerical covariates.

Table 3.2 Shift in weight category from baseline while on AAP treatment.

	Underweight on AAP	Normal on AAP	Overweight on AAP	Obese on AAP
Underweight (< 18.5)	9 (75%)	3 (25%)	0 (0%)	0 (0%)
Normal (18.5-24.9)	10 (3.4%)	239 (81%)	42 (14%)	3 (1%)
Overweight (25-29.9)	0 (0%)	59 (20%)	209 (71%)	26 (8.9%)
Obese (>30)	0 (0%)	1 (0.45%)	28 (13%)	195 (87%)

Figure 3.2 Change in BMI histogram.

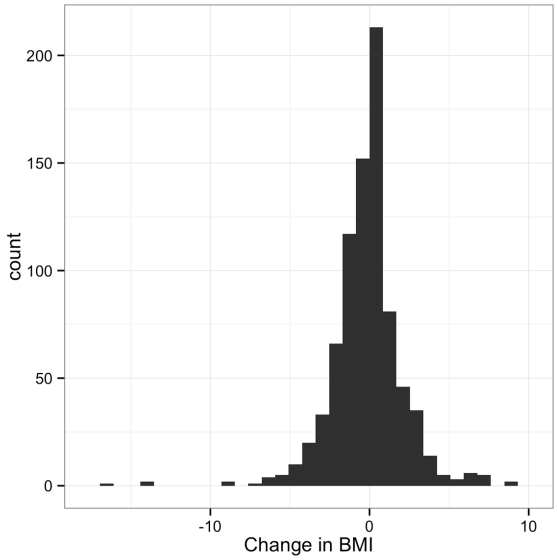
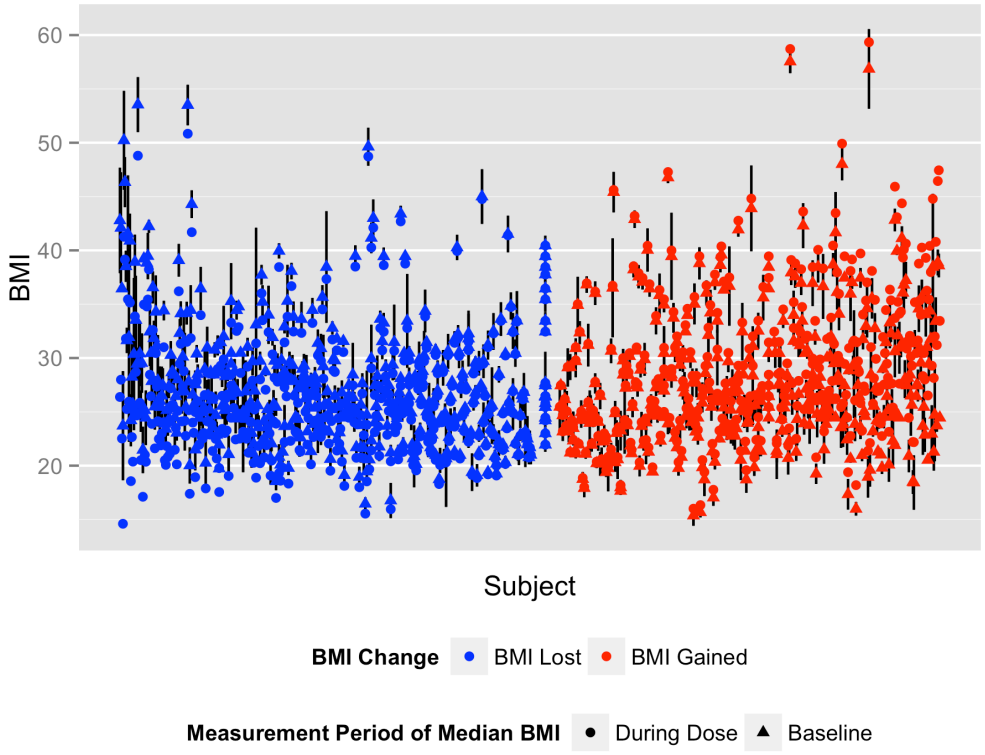
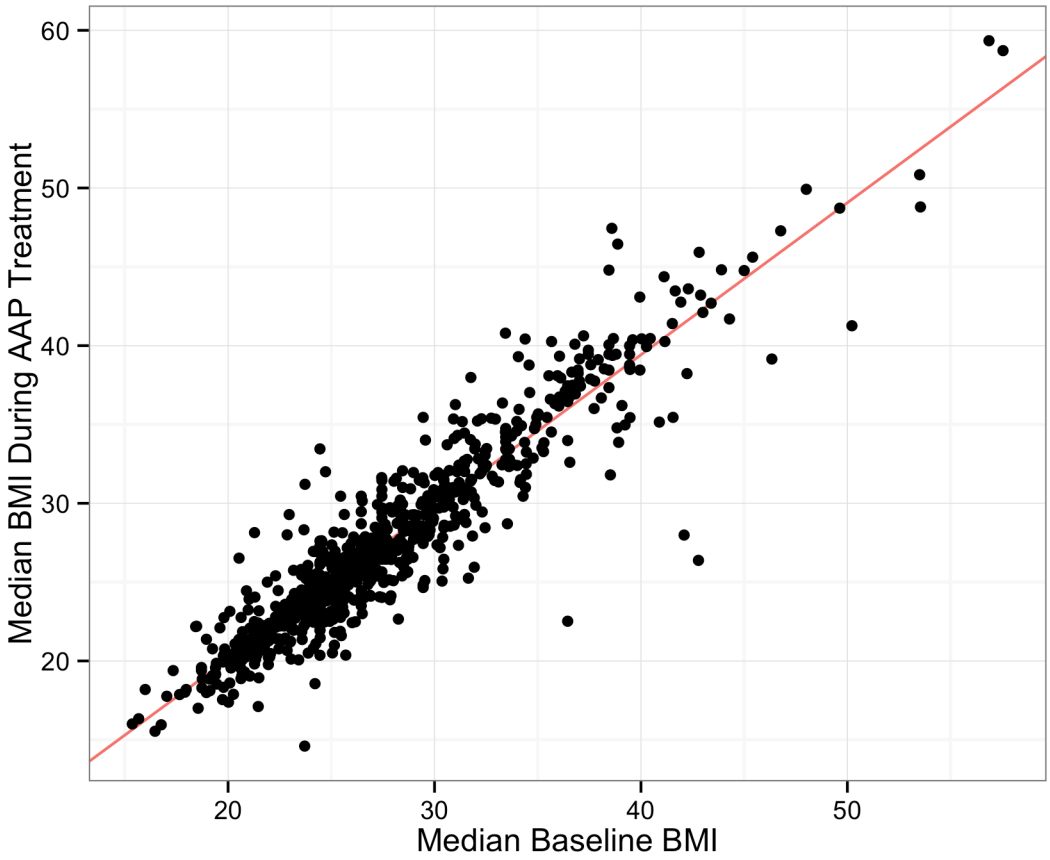


Figure 3.3 BMI differences between baseline and treatment measurement periods.



Each vertical line represents the BMI range of a single subject over both baseline and treatment measurement periods. Circles represent a subject's median BMI during AAP treatment, while triangles represent their median baseline BMI. Subjects who lost weight while on an AAP are shown in blue, while those who gained weight are shown in red.

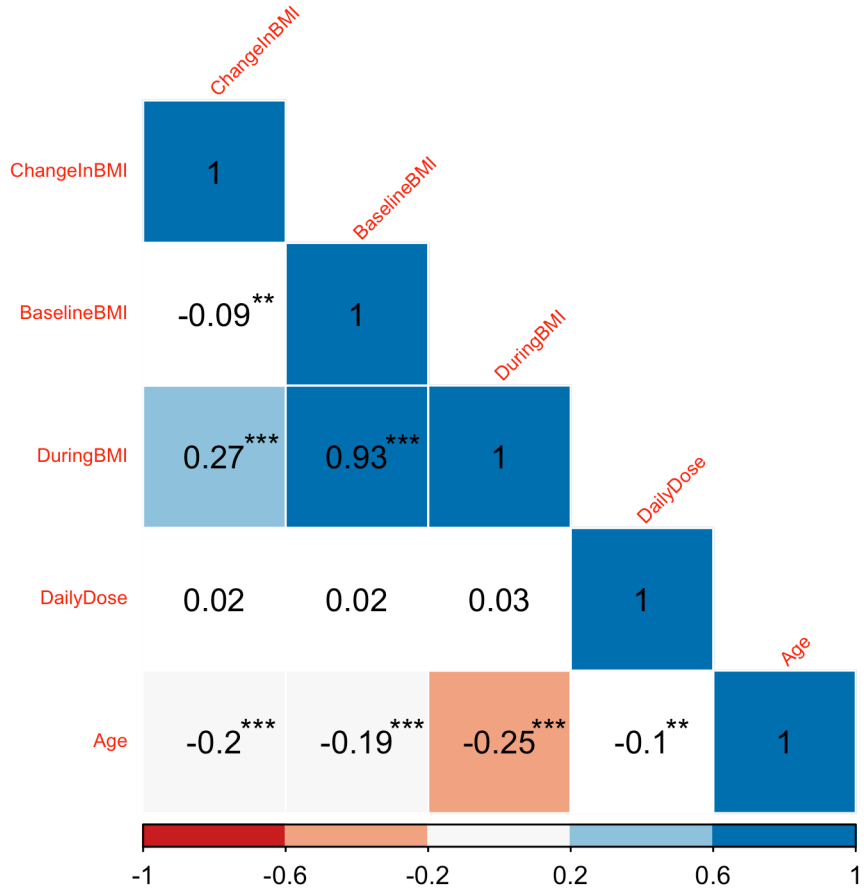
Figure 3.4 Median baseline BMI versus median BMI during AAP treatment.



3.2.4 GWAS

The two extracted phenotypes, change in BMI and greater than 7% increase in BMI, were used as outcome values for two GWAS spanning 567,096 SNPs after genotype quality control steps. Cohort characterization and covariates for the 823 subjects are shown in Table 3.1. GWAS of the greater than 7% increase from baseline BMI dichotomous phenotype did not yield significant results at $P < 5*10^{-8}$ significance cutoff.

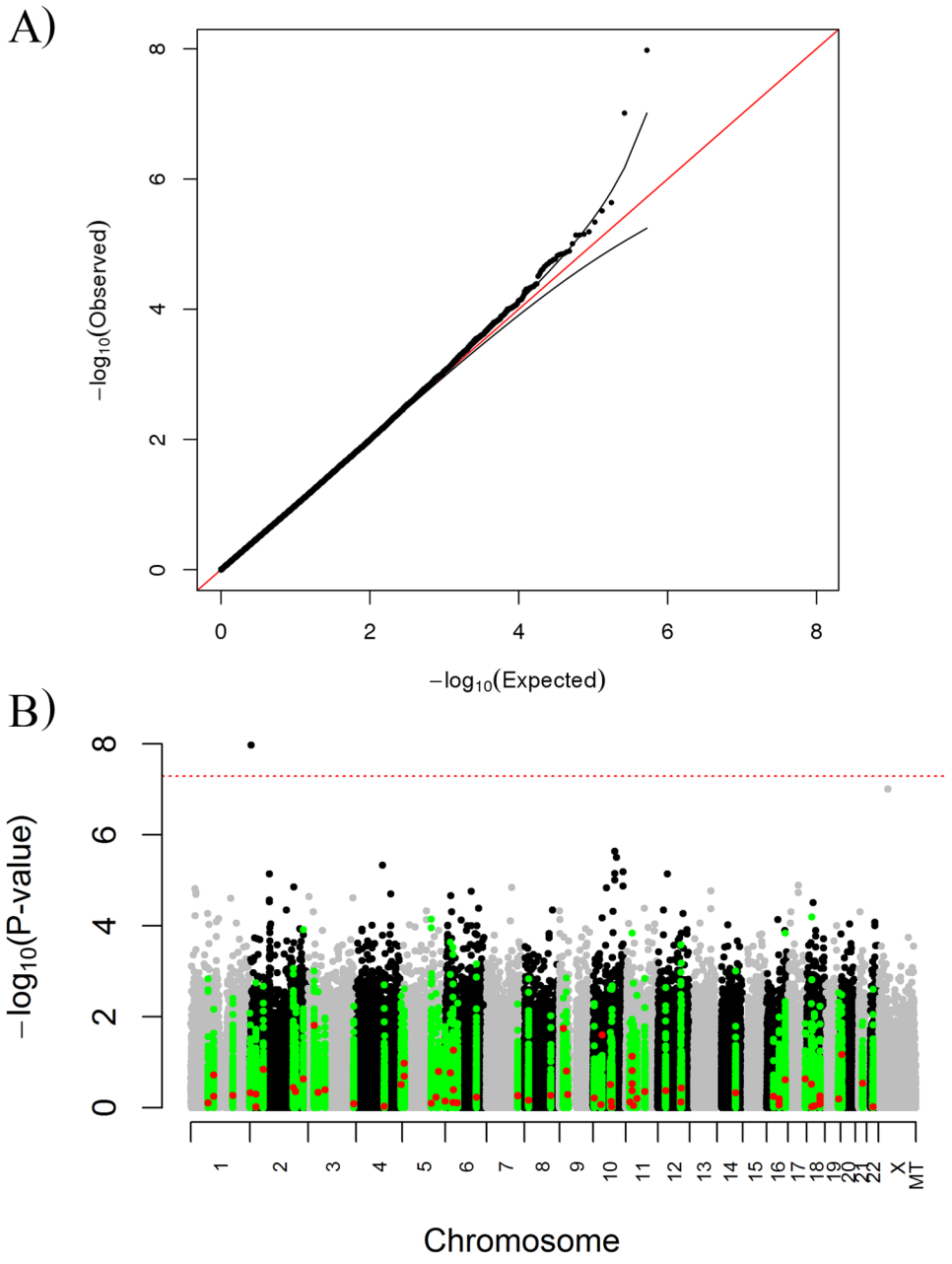
Figure 3.5 Correlation matrix plot of change in BMI, baseline BMI, BMI during treatment, daily AAP dose, and age.



Correlation plot of BMI phenotypes and other numerical covariates. Values represent Pearson's correlation coefficients, ranging from -1 (red) to 1 (blue). Significant correlations after multiple testing correction are indicated as follows: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

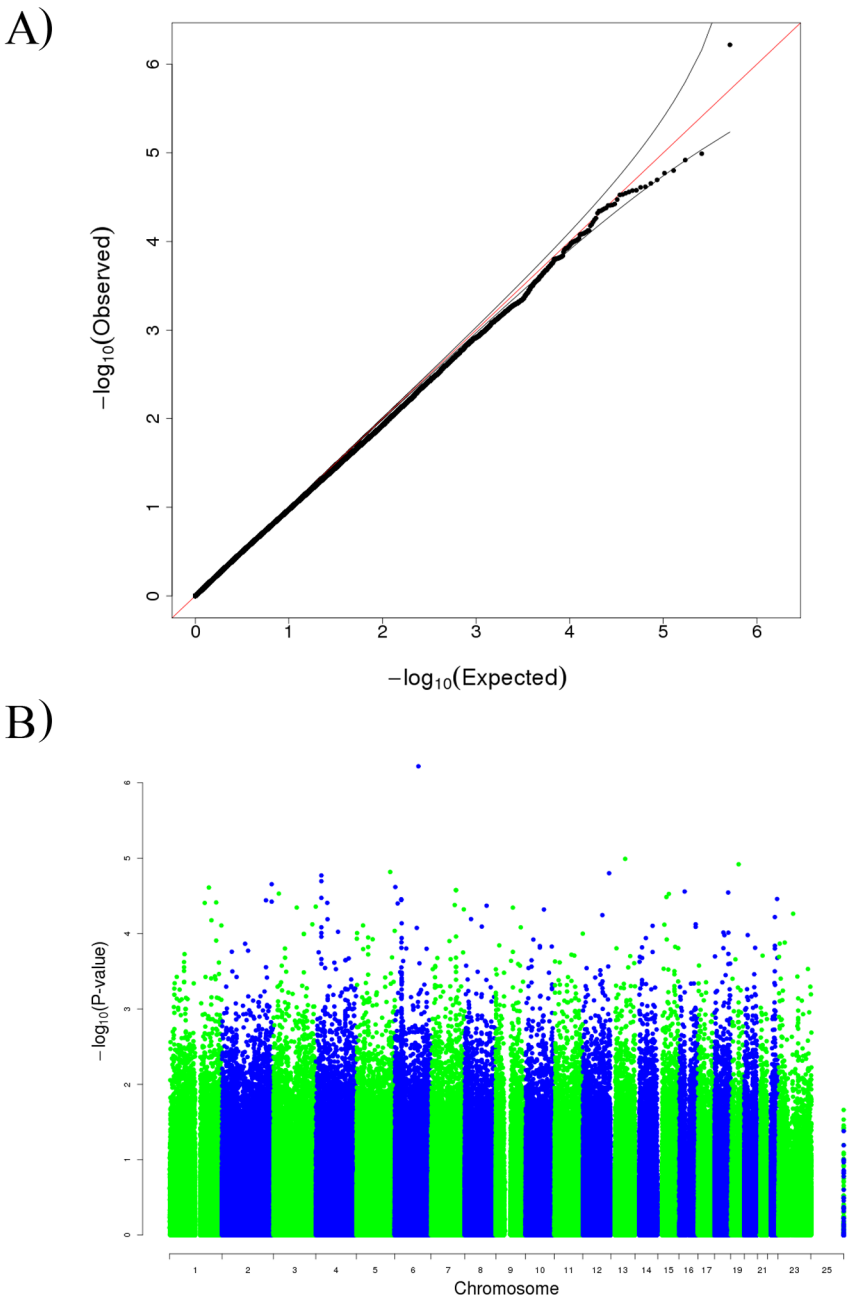
GWAS of the change in BMI phenotype yielded one significant SNP (rs80167927) in an intergenic region of chromosome 2 with a p-value of $1.05*10^{-8}$ and beta of -2.0. A Q-Q plot and Manhattan plot of the change in BMI phenotype is shown in Figure 3.6, while Figure 3.7 shows similar figures for the greater than 7% increase from baseline BMI phenotype. A list of the top associated SNPs ($P < 10^{-5}$) for the change in BMI GWAS is shown in Table 3.3. 3 SNPs (rs11202805, rs12356091, and rs7083493) are in strong linkage disequilibrium (D'=1) with rs6586145, a missense variant in the gastric lipase gene *LIPF*.

Figure 3.6 GWAS results for change in BMI phenotype.



A) Q-Q plot of change in BMI with 95% confidence intervals. The distribution has a genomic inflation factor λ_{GC} of 0.98. B) Manhattan plot for GWAS of change in BMI. Red dots indicate a candidate SNP for obesity or BMI, as reported in the NHGRI GWAS catalog. Green dots indicate SNPs within 1 MB of a candidate SNP.

Figure 3.7 GWAS results for greater than 7% increase in baseline BMI phenotype.



A) Q-Q plot of greater than 7% increase in baseline BMI phenotype with 95% confidence intervals. The distribution has a genomic inflation factor λ_{GC} of 1.1. B) Manhattan plot for GWAS of greater than 7% increase in baseline BMI, colored by chromosome.

Table 3.3 Top associated SNPs for change in BMI GWAS

SNP	Major Allele	Minor Allele	Chr	Beta	P-Value	BMI Gained MAF % (Count)	BMI Lost MAF % (Count)	Chip MAF (%)	Gene	Gene Location
rs80167927	T	C	2	-2	$1.05*10^{-8}$	1.4 (12)	3.2 (26)	2.7		
rs6610297	C	T	X	-3.1	$9.74*10^{-8}$	0.62 (5)	1.4 (12)	1.1		
rs11202805	C	T	10	-0.9	$2.30*10^{-6}$	6.9 (57)	10.4 (86)	7.8	LIPF	intron
rs1326830	C	A	10	-2.8	$3.09*10^{-6}$	0.40 (3)	1.2 (10)	0.8	CYP2C18	3'UTR
rs1967654	C	T	10	0.53	$6.47*10^{-6}$	32.7 (269)	26.7 (220)	29.4	DMBT1	intron
rs12356091	G	A	10	-0.9	$7.04*10^{-6}$	6.8 (56)	10.9 (90)	7.9	LIPK	intron
rs59787038	A	G	2	-0.6	$7.25*10^{-6}$	25.1 (207)	29.8 (245)	27.9		
rs12049977	T	C	12	-1.3	$7.27*10^{-6}$	3.2 (26)	4.5 (37)	4.3		
rs7083493	C	T	10	-0.8	$9.86*10^{-6}$	7.0 (58)	11.7 (96)	8.2		

SNPs with p-values less than $1.0x10^{-5}$ shown. Chr: chromosome, MAF: Minor Allele Frequency.

Using both phenotypes, we then performed a post-hoc candidate SNP assessment using a list of 73 SNPs from 43 genes identified from literature as being associated with BMI, obesity, or AAP drug response. Results of these SNPs for the change in BMI phenotype are colored in red in the Manhattan plot in Figure 3.6B, with green indicating SNPs within a 1 Mb region surrounding each candidate SNP. However, none of these SNPs were statistically significant following multiple testing correction for either phenotype. Candidate SNPs with unadjusted p-values < 0.05 for either phenotype are reported in Table 3.4.

Table 3.4 Post-hoc candidate SNP analysis results.

SNP	Chr	Beta / Odds Ratio	p-value	Gene	Location	Association	Phenotype
rs6787891	3	0.56	0.015	RARB	intron	BMI	Change in BMI
rs1927702	9	0.26	0.018			BMI	Change in BMI
rs7474896	10	-0.37	0.030			Obesity	Change in BMI
rs10458787	10	1.73	0.049			BMI	>7% Change in BMI
rs12419692	11	0.52	0.030			LD with BMI SNP	>7% Change in BMI

Results of post-hoc candidate SNP analysis. While no SNP was statistically significant after multiple testing correction, SNPs with unadjusted p-value < 0.05 are shown here. Candidate SNPs were derived from NHGRI GWAS Catalog SNPs for obesity or BMI phenotypes as listed in "Association" column. "Phenotype" column refers to the GWAS phenotype where the result was derived.

3.3 Materials & Methods

3.3.1 Study Population

Subjects in this study are part of the Kaiser Permanente Research Program on Genes, Environment and Health (RPGEH) Genetic Epidemiology Research on Adult Health and Aging (GERA, N=102,979) cohort. A detailed description of the GERA cohort has been previously

described in Chapter 2.3.1. For sample homogeneity, this study was restricted to cohort members who self-identified as Non-Hispanic Whites and were run on the European genotyping array.

3.3.2 Genotype Data

As described in Chapter 2.3.2, Genotyping for the GERA cohort was performed using four custom Affymetrix Axiom arrays created for the four major race-ethnicity groups in the cohort: African Americans (AFR), East Asians (EAS), Latinos (LAT), and Non-Hispanic Whites (EUR)[14,15]. For this study, only Non-Hispanic White individuals run on the EUR arrays were included. Principal components analysis was performed using Eigenstrat[17] to assess population substructure in the study samples. The first three principal components explained most variation and were used in the GWAS to characterize genetic ancestry.

3.3.3 Prescription Data

Drug identifiers were obtained from dispensed outpatient pharmacy records for all drugs with an active ingredient matching “quetiapine”, “aripiprazole”, “olanzapine”, “clozapine”, “risperidone”, “ziprasidone”, “paliperidone”, or “iloperidone” in the generic label name, and drug class “antipsychotics”. Records were filtered to only include those dispensed and not returned to stock. Prescription records were available in 1995, and were queried up to and including January 1, 2012. Daily doses were calculated as dose multiplied by frequency per day. Any prescriptions with unknown frequency or total duration less than 30 days were removed. If there was a gap of less than 28 days between the end of one prescription and the start of another, the two were combined to be one single prescription.

3.3.4 Phenotype

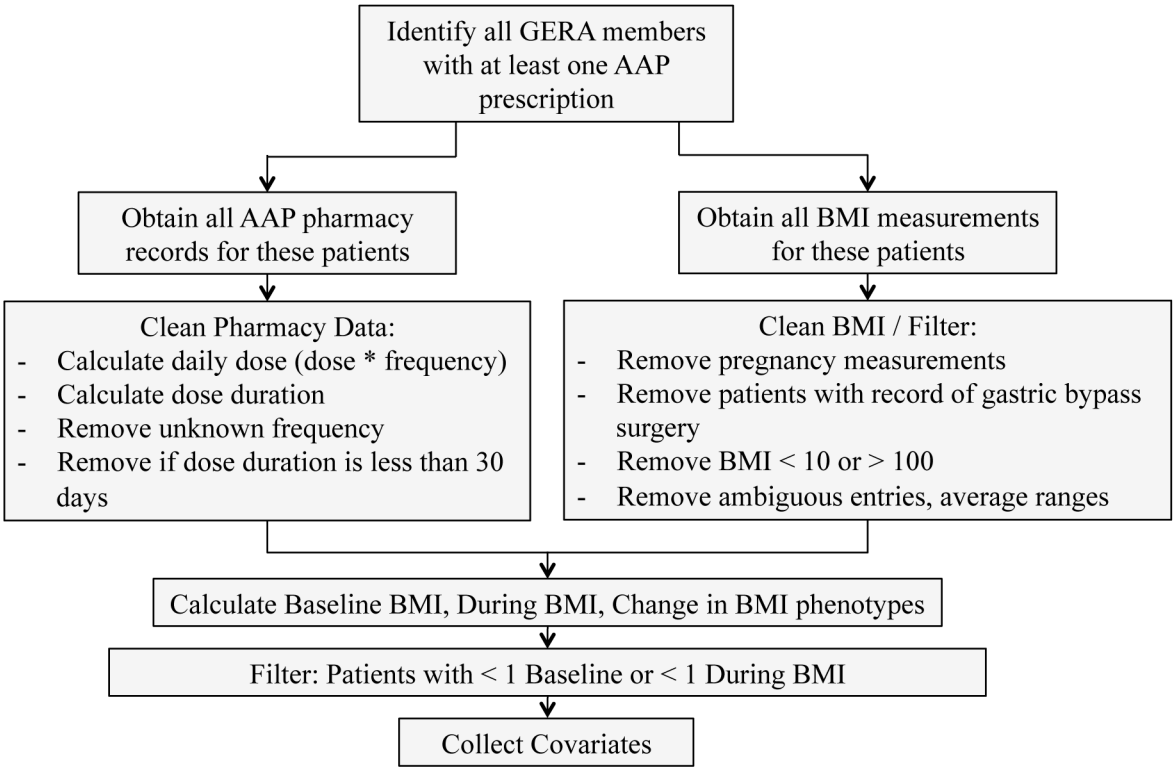
For all GERA members with at least one valid prescription record for an AAP, weight data was extracted from Vitals tables in the EHR system. Weight records varied in format,

generally showing weight in pounds or kg at time of appointment, or in BMI. If weight was not recorded in BMI format, it was calculated as weight (kg) / height $(m)^2$ using the closest available height measurement. Subjects were excluded if BMI measurements could not be calculated due to lack of height data.

BMI measurements were filtered to exclude any measurements taken during pregnancy or emergency room visits. Subjects that had undergone gastric bypass procedures were removed from the cohort. Any ambiguous BMI entries (e.g. "BMI < 18.0") or erroneous entries with BMI less than 10 or greater than 100 were removed. For BMI measurements that were recorded as ranges (e.g. "24-24.9"), mean values were calculated. Figure 3.8 shows an overview of the data collection and cleaning process.

Change in BMI phenotypes were calculated according to the schematic in Figure 3.1. We identified the first AAP prescription where a BMI measurement was taken during the dose duration. The subject was identified as AAP naïve if the first dose with a BMI measurement available was the subject's first time receiving an AAP. They were identified as not AAP naïve if this was not the case, that is, if they received an earlier AAP prescription but no BMI measurements were taken during that dosage period. A baseline BMI was calculated by taking the median BMI measurement over the 3 years prior to the first AAP prescription. A BMI measurement was calculated to represent the subject's BMI while on the AAP by taking the median BMI over all measurements recorded during the dosage period. A change in BMI was calculated as the difference between the BMI measurement during AAP use and the baseline BMI. We also calculated an additional phenotype to identify subjects who gained or lost greater than 7% of their baseline BMI.

Figure 3.8 Phenotype calculation workflow diagram.



Covariates for GWAS included age at the first AAP dose with BMI data available, sex, genetic ancestry principal components, mean baseline BMI, duration of first dose with BMI data available, daily dose for first AAP dose, whether the subject was AAP naïve, type 2 diabetes status, smoking status, and AAP name.

3.3.5 Analysis

Data cleaning, processing, and phenotype extraction and analysis was performed in R version 2.15[18]. Genome-wide association analyses were performed in PLINK v1.08[19] using a linear regression additive model for the change in BMI phenotype, or a logistic regression additive model for greater than 7% increase from baseline BMI phenotype. Linkage disequilibrium analysis was performed using SNAP[20]. Follow-up candidate SNPs assessment

was performed for SNPs associated with obesity or BMI phenotypes in the NHGRI GWAS Catalog[21].

3.4 Discussion

This study presents the largest analysis to date of clinically observed AIWG, a phenotype that is typically difficult and costly to measure. Using data extracted from an EHR, we observed similar frequencies of this side effect as has been reported in clinical datasets and controlled trials[22-24]. While only one SNP reached genome-wide statistical significance for association with change in BMI, we found a number of modestly associated SNPs that have not been previously reported and may warrant further investigation.

The top associated SNP in our analysis of change in BMI was rs80167927, an intergenic SNP on chromosome 2. This variant has a 5% minor allele frequency in 1000 Genomes, and appeared in 26 patients who lost weight while on AAP and 12 patients who gained weight. Further assessment of this SNP did not yield significant biological findings, as it is not located near a gene related to the phenotype, nor is it in a regulatory region[25].

While other SNPs failed to reach genome-wide statistical significance in this analysis, a few modestly associated SNPs ($P < 10^{-5}$) were found in genes that may be related to weight loss. 3 SNPs (rs11202805, rs12356091, and rs7083493) were in strong linkage disequilibrium with a missense variant in *LIPF*, which encodes gastric lipase, a protein involved in the digestion of dietary triglycerides in the gastrointestinal tract. The minor alleles of these 3 SNPs were present in more subjects that had lost BMI (10%, 11%, and 12%, respectively, of those who lost weight), with beta values ranging from -0.9 to -0.8, suggesting implications in weight loss. One of the three SNPs, rs12356091, is located in an intron of *LIPK*, which plays a role in keratinocyte differentiation in epidermal layers. A post-hoc analysis of candidate SNPs from BMI- and AAP-

related pathways did not reveal significant findings after adjusting for multiple testing. Notably, variants in candidate genes like *HTR2C* and *MC4R* that have previously been reported as associated with AIWG did not replicate in this study.

There were major challenges in capturing drug response data from this resource. For one, patient compliance, which is likely to decrease if the patient perceives that the drug is causing them to gain weight, is difficult to capture from EHR data. As a proxy, we looked at factors like whether a prescription was being consistently picked up (not returned to stock), and if the patient had at least one visit with a physician during the time they were on the AAP. Additionally, we observed a great deal of variability in phenotype, timing, and number of measurements among subjects. Weight was reported in various formats and may have been recorded using varying tools. For example, a manual scale may be more prone to measurement error than a digital one, and some healthcare providers may adjust for clothing weight while others do not, although such variability may be reduced with multiple measurements. There also may have been a number of clinical factors that we did not adjust for, such as comorbidities and drug interactions.

One major limitation of this study was restriction of subjects by ethnicity. To attempt to replicate findings associated with AIWG in the literature, we sought to assemble the largest, most homogeneous cohort that was representative of cohorts in those studies. This included Non-Hispanic Whites, which made up the majority of the GERA cohort. To increase sample size, a follow-up study may include a meta-analysis of all AAP users with BMI information, regardless of ethnicity.

Additionally, our analysis was inclusive of all AAPs. As a drug class, more subjects in our cohort lost weight while on AAPs than gained it. This may have been due to the proportion of drugs with different risks for weight gain. For example, olanzapine and clozapine, the AAPs

with the highest risk of weight gain[2], were only taken by 12% of the cohort. Also, our analysis did not adjust for indication. The genetic variants associated with response to AAPs may differ in schizophrenic patients compared to patients with bipolar disorder, for example. We also observed that some AAPs were often prescribed at a much lower dose than used for psychiatric disorders, and may have been used off-label for indications such as insomnia, which is difficult to capture from EHR data. A stratified analysis or a study restricted to a single indication may yield more significant results.

While 65% of our cohort had only been prescribed a single AAP, the remaining 35% may be of interest for a follow-up study to assess the impacts of switching AAPs. AIWG is reported to occur within 6 months of starting therapy[26], and may be inversely correlated with treatment efficacy[24]. By extending the follow-up period and collecting more data on treatment response, this resource may be useful for studying the long-term effects of AAPs. Additionally, including more data on other metabolic phenotypes, such as lipid or blood glucose levels, will help form a better picture of the overall effects of AAP use and potentially yield novel genetic findings.

3.5 References

1. Leucht, S. *et al.* Second-Generation Versus First-Generation Antipsychotic Drugs For Schizophrenia: A Meta-Analysis. *Lancet* **373,** 31–41 (2009).
2. Müller, D. J. & Kennedy, J. L. Genetics Of Antipsychotic Treatment Emergent Weight Gain In Schizophrenia. *Pharmacogenomics* **7,** 863–887 (2006).
3. Ascher-Svanum, H., Stensland, M. D., Kinon, B. J. & Tollefson, G. D. Weight Gain As A Prognostic Indicator Of Therapeutic Improvement During Acute Treatment Of Schizophrenia With Placebo Or Active Antipsychotic. *J Psychopharmacol* **19**, 110-117 (2005).

4. Basson, B. R., Kinon, B. J. & Taylor, C. C. Factors Influencing Acute Weight Change In Patients With Schizophrenia Treated With Olanzapine, Haloperidol, Or Risperidone. *J Clin Psychiatry* **62**, 231-238 (2001).

5. Seeman, M. V. Secondary Effects of Antipsychotics: Women at Greater Risk Than Men. *Schizophr Bull* **35,** 937–948 (2009).

6. Shams, T. A. & Müller, D. J. Antipsychotic Induced Weight Gain: Genetics, Epigenetics, And Biomarkers Reviewed. *Curr Psychiatry Rep* **16,** 473 (2014).

7. Lieberman, J. A. *et al.* Effectiveness Of Antipsychotic Drugs In Patients With Chronic Schizophrenia. *N Engl J Med* **353,** 1209–1223 (2005).

8. Need, A. C. *et al.* Pharmacogenetics Of Antipsychotic Response In The CATIE Trial: A Candidate Gene Analysis. *Eur J Hum Genet* **17,** 946–957 (2009).

9. Tiwari, H. K. *et al.* Association Of Allelic Variation In Genes Mediating Aspects Of Energy Homeostasis With Weight Gain During Administration Of Antipsychotic Drugs (CATIE Study). *Front Genet* **2**, 56 (2011). doi: 10.3389/fgene.2011.00056

10. Adkins, D. E. *et al.* Genomewide Pharmacogenomic Study Of Metabolic Side Effects To Antipsychotic Drugs. *Mol Psychiatry* **16,** 321–332 (2011).

11. Malhotra, A. K. *et al.* Association Between Common Variants Near The Melanocortin 4 Receptor Gene And Severe Antipsychotic Drug-Induced Weight Gain. *Arch Gen Psychiatry* **69,** 904–912 (2012).

12. Chowdhury, N. I. *et al.* Genetic Association Study Between Antipsychotic-Induced Weight Gain And The Melanocortin-4 Receptor Gene. *Pharmacogenomics J* **13,** 272–279 (2013).

13. Kvale, M. N. *et al.* Genotyping Informatics and Quality Control for 100,000 Subjects in

the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200,** 1051–1060 (2015).

14. Hoffmann, T. J. *et al.* Design And Coverage Of High Throughput Genotyping Arrays Optimized For Individuals Of East Asian, African American, And Latino Race/Ethnicity Using Imputation And A Novel Hybrid SNP Selection Algorithm. *Genomics* **98,** 422–430 (2011).

15. Hoffmann, T. J. *et al.* Next Generation Genome-Wide Association Tool: Design And Coverage Of A High-Throughput European-Optimized SNP Array. *Genomics* **98,** 79–89 (2011).

16. Banda, Y. *et al.* Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200,** 1285–1295 (2015).

17. Price, A. L. *et al.* Principal Components Analysis Corrects For Stratification In Genome-Wide Association Studies. *Nat. Genet.* **38,** 904–909 (2006).

18. R Core Team. R: A Language And Environment For Statistical Computing. R Foundation For Statistical Computing. http://www.r-project.org/ (2015).

19. Purcell, S. *et al*. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* **81**, 559-575 (2007).

20. Johnson, A. D. *et al.* SNAP: A Web-Based Tool For Identification And Annotation Of Proxy Snps Using Hapmap. *Bioinformatics* **24,** 2938–2939 (2008).

21. Welter, D. *et al.* The NHGRI GWAS Catalog, A Curated Resource Of SNP-Trait Associations. *Nucleic Acids Res* **42,** D1001–1006 (2014).

22. Bushe, C. J., Slooff, C. J., Haddad, P. M. & Karagianis, J. L. Weight Change From 3-Year

Observational Data: Findings From The Worldwide Schizophrenia Outpatient Health Outcomes Database. *J Clin Psychiatry* **73,** e749–755 (2012).

23. Cohen, D., Bonnot, O., Bodeau, N., Consoli, A. & Laurent, C. Adverse Effects Of Second-Generation Antipsychotics In Children And Adolescents: A Bayesian Meta-Analysis. *J Clin Psychopharmacol* **32,** 309–316 (2012).

24. Hermes, E. *et al.* The Association Between Weight Change And Symptom Reduction In The CATIE Schizophrenia Trial. *Schizophr Res.* **128,** 166–170 (2011).

25. Boyle, A. P. *et al.* Annotation Of Functional Variation In Personal Genomes Using Regulomedb. *Genome Research* **22,** 1790–1797 (2012).

26. Panariello, F., Polsinelli, G., Borlido, C., Monda, M. & De Luca, V. The Role Of Leptin In Antipsychotic-Induced Weight Gain: Genetic And Non-Genetic Factors. *J Obes* **2012,** 572848 (2012).

4 EHR-DERIVED DRUG RESPONSE: CLOPIDOGREL AND MAJOR ADVERSE CARDIOVASCULAR EVENTS

4.1 Introduction

Clopidogrel, the focus of this chapter, is a widely prescribed antiplatelet drug used to reduce ischemic complications in a number of patients with coronary artery disease. Clopidogrel is a thienopyridine used to treat unstable angina and ST-segment elevation myocardial infarction (MI), as well as stroke or established peripheral arterial disease. For those undergoing percutaneous coronary intervention (PCI), clopidogrel plus aspirin remain standard of care. With the expiration of the Plavix patent in 2012 and the FDA approval of generic manufacturing[1], clopidogrel will continue to be a widely prescribed antiplatelet therapy. An estimated 20-50% of patients that do not respond to clopidogrel[2-4] (depending on non-responsiveness definition) may experience a severe event, such as recurrent myocardial infarction, bleeding, neutropenia, or gastrointestinal problems. Thus quantifying and predicting clopidogrel response has been a major focus for the pharmacogenomics research community.

One primary cause of variability in clopidogrel response is deficiencies in genes related to metabolic activation of clopidogrel. Clopidogrel is a prodrug, with numerous CYP450 enzymes involved in its conversion to the active metabolite, including CYP2C19, CYP1A2, CYP2B6, and CYP3A4/5[5,6]. Only 15% of clopidogrel gets converted to its active metabolite; the rest being excreted[7]. The active metabolite then irreversibly binds to the P2Y12 receptor of platelets, blocking ADP-binding and reducing platelet activation and subsequent aggregation.

While earlier studies considered variants in clopidogrel's target P2Y12[8-10], variation in clopidogrel response could not be consistently explained until the mid-2000s, when studies began focusing on metabolism of clopidogrel. Candidate gene studies focused on the CYP450 genes involved in clopidogrel bioactivation, with many key studies appearing in 2009 tying

variants in *CYP2C19* with increased risk of cardiovascular events following stent placement and clopidogrel therapy[2,11,12]. One genome-wide association study (GWAS) of clopidogrel response was performed on a cohort of healthy Amish subjects, also identifying *CYP2C19* as the locus that explains the most variation in response[13]. Various ongoing studies seek to understand clopidogrel response in larger and more diverse populations of patients.

Defining clopidogrel non-responsiveness is non-trivial, with different definitions making comparisons between studies difficult[3]. While clopidogrel platelet aggregation tests provide a direct measure of clopidogrel response, the test itself is time-sensitive, as it must be performed, with results returned, prior to a PCI procedure. A cutoff of laboratory values of platelet aggregation can then be used to define clopidogrel responsiveness. Alternatively, a proxy phenotype can be used, such as hard outcomes of major cardiovascular events during or following therapy. In addition to differences in drug response definition, study populations vary, either in terms of ethnicity or prior disease or treatment status, making replication efforts difficult. Nevertheless, in 2010, the FDA recognized that variants in *CYP2C19* appear to be the primary genetic factor influencing clopidogrel non-responsiveness, changing the drug label to include a Boxed Warning[14]. In 2011, the Clinical Pharmacogenetics Implementation Consortium (CPIC) published guidelines on *CYP2C19* genotype-directed antiplatelet therapy based on available literature, providing suggestions for which patients to test and how to clinically act on genotyping results[4].

In this chapter we utilize electronic health records (EHR) from the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) to examine clopiodgrel response in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. We

utilize *CYP2C19* metabolizer status phenotypes generated in Chapter 2 to assess the impact of clinically actionable pharmacogenetic information on this drug response phenotype.

4.2 Results

4.2.1 Cohort Description

6617 GERA cohort members had at least one outpatient prescription of clopidogrel. Of these, 1723 had a record of MI followed by a clopidogrel prescription within 30 days of the MI. 745 cohort members had a record of a PCI procedure with a drug-eluting stent, followed by a clopiodgrel prescription within 30 days of the procedure. 358 had an MI followed by a PCI procedure, along with a clopidogrel prescription within 30 days. 2110 individuals total had either or both initial events.

Records of major adverse cardiovascular events (MACE) were obtained for these individuals. 419 had evidence of MACE within the 730 days following the first clopidogrel prescription after the initial event, and are considered non-responders of clopidogrel in this study. 435 individuals had MACE occur after 730 days, and were excluded from the analysis. 1256 responders total did not have any record of MACE following the first clopidogrel prescription. A description of the cohort of 1675 individuals included in the analysis is shown in Table 4.1.

4.2.2 Phenotype

For the 419 non-responders, time to MACE in days was calculated and is shown in Figure 4.1. For 73% of non-responders (N=305), MACE occurred within 30 days of the initial clopiodgrel prescription. 94% of non-responders (N=392) had MACE occur within 1 year. MACE occurrence differed significantly by initial event ($\chi^2(2) = 10.3$, $P < 0.01$). A greater percentage of those who had MI only as an initial event experienced MACE (27%), compared to

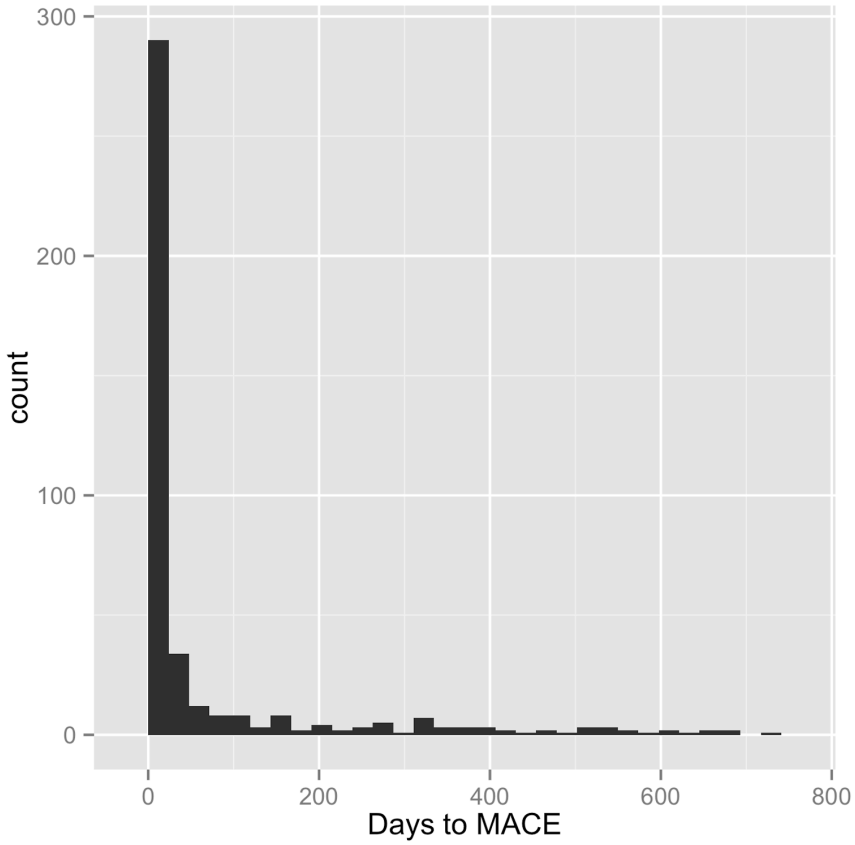
those who only had PCI but no MACE (19%), or those who had MI followed by PCI (23%). A significantly greater percentage of females (29%) experienced MACE compared to males (24%) (χ^2 (1) = 4.5, $P < 0.05$). MACE occurrence did not vary significantly by CYP2C19 metabolizer status ($\chi^2(2) = 3.7$, $P = 0.16$), CYP2C19 diplotype ($\chi^2(4) = 4.7$, $P = 0.32$; exclusive of CYP2C19*3 haplotypes due to low sample size), smoking status ($\chi^2(1) = 0.33$, $P = 0.56$), or race ($\chi^2(4) = 2.1$, $P = 0.73$). Time to MACE was not significantly correlated with age at initial event ($r = 0.05$, $P = 0.31$).

Table 4.1 Cohort description

	N	Percentage
Total	1675	--
Male	1178	70.33%
Non-Hispanic White	1426	85.13%
African American	38	2.24%
Asian	100	5.97%
Latino	101	6.03%
Other/Uncertain	10	0.60%
Smoker	881	52.60%
Initial event: MI only	1090	65.07%
Initial event: PCI only	305	18.21%
Initial event: MI, then PCI	280	16.72%
MACE occurrence within 30 days	305	18.21%
MACE occurrence within 365 days	392	23.40%
MACE occurrence within 730 days	419	25.01%
	Mean	SD
Age at Initial Event (years)	70.4	9.9
Clopiodgrel Exposure (days, up to 730 days)	246	222

Description of the 1675 cohort members in this study. Smoking status is defined as self-reported current or former smoker in lifestyle surveys conducted by the RPGEH. Clopidogrel exposure indicates the cumulative number of days that each subject had a valid clopiodgrel prescription, regardless of whether MACE occurred, in the 730 day period following the first clopidogrel prescription after the initial event.

Figure 4.1 Histogram of time to MACE for clopidogrel non-responders



A Cox regression analysis was performed based on time to MACE, with a time-dependent covariate of cumulative clopidogrel exposure. The assumption of proportionality of hazards was valid for this model, with global $\chi^2(11) = 0.45$, $P=1$. We found that those who had only an initial MI (HR = 2.48, 95% CI=2.36-2.61, P<0.001) experienced significantly more MACE than those who had a PCI only or both a PCI and MI. African Americans experienced significantly more MACE than any other ethnicity (HR = 2.42, CI = 2.16-2.71, P<0.001). Smokers had a slightly greater probability of MACE than non-smokers (HR = 1.25, 95% CI = 1.22-1.28, P<0.001). Kaplan-Meier curves stratified by initial event, ethnicity, and CYP2C19 metabolizer status are shown in Figure 4.2 to 4.4.

Figure 4.2 Kaplan-Meier curve stratified by initial event

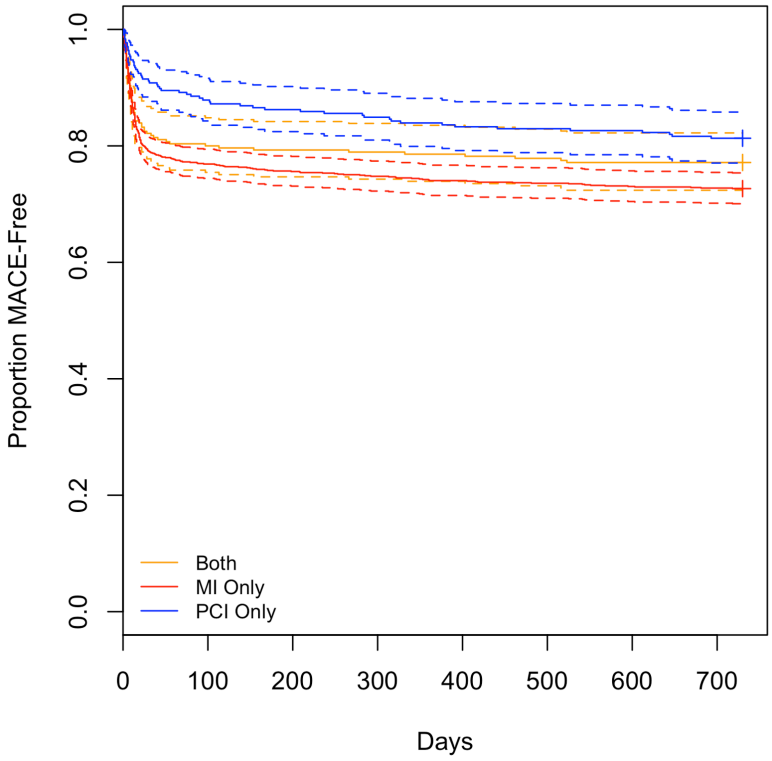


Figure 4.3 Kaplan-Meier curve stratified by ethnicity

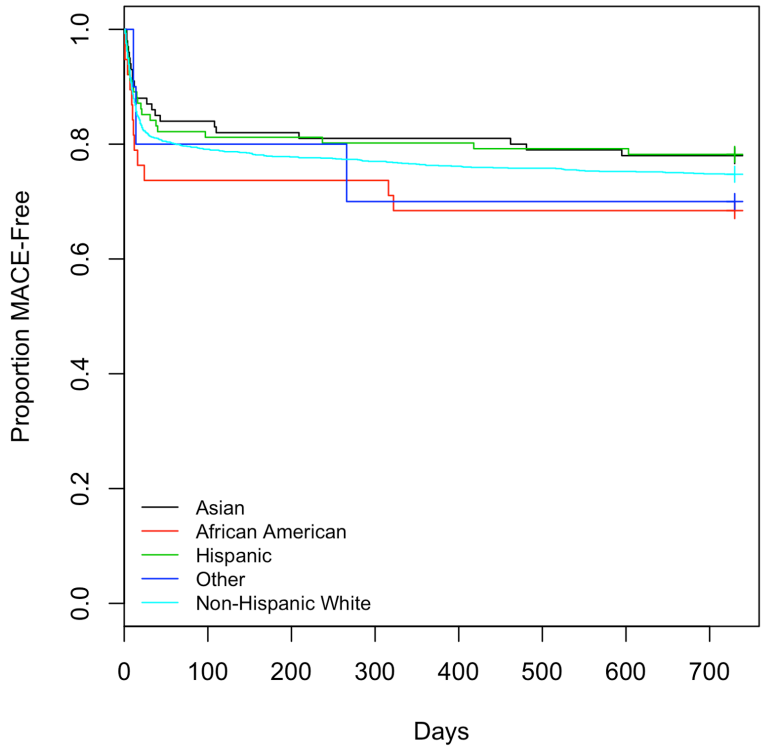
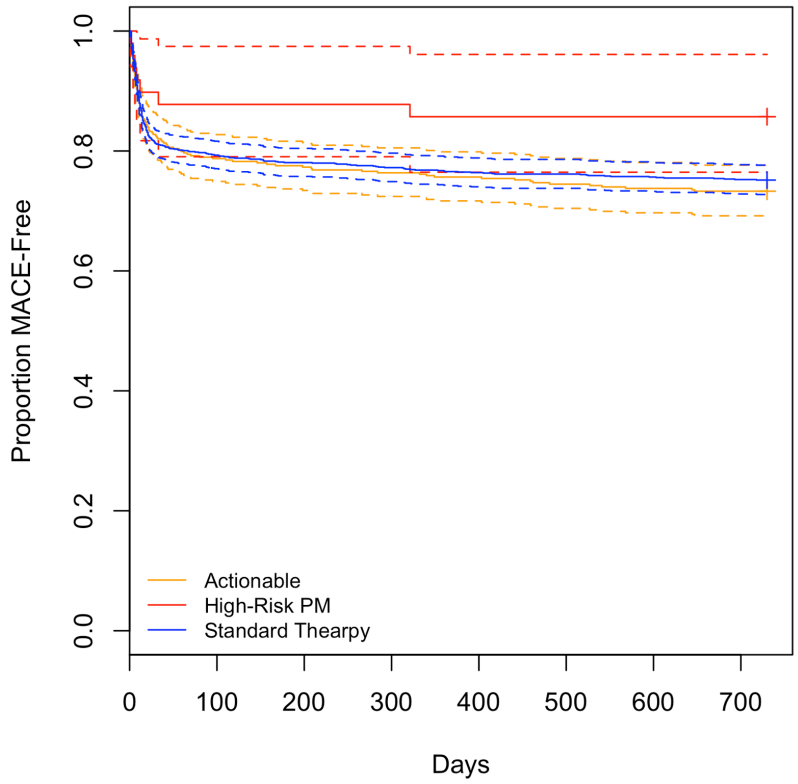
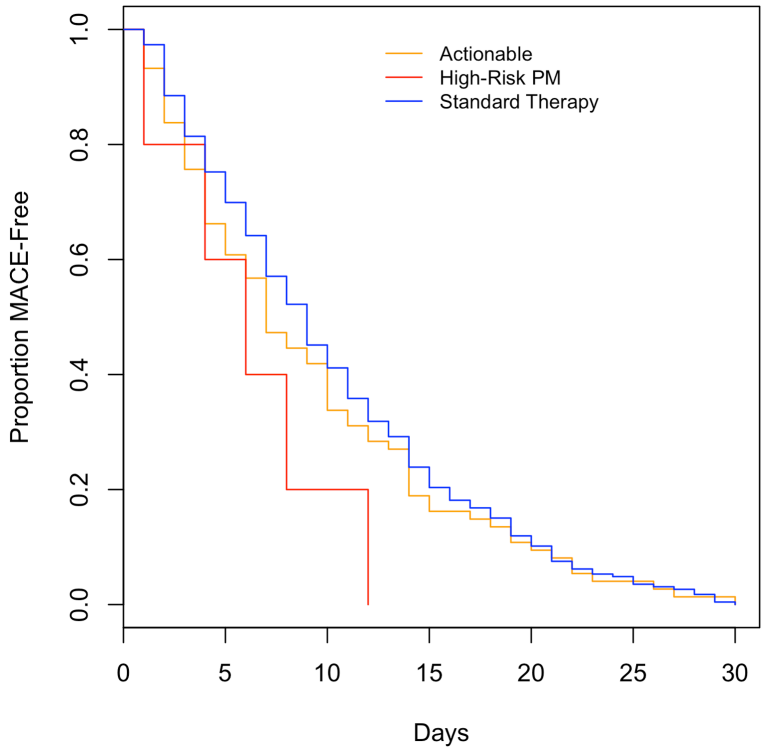


Figure 4.4 Kaplan-Meier curve stratified by CYP2C19 metabolizer status



Unexpectedly, CYP2C19 metabolizer status was not predictive of MACE in the cohort. In fact, high-risk poor metabolizers were significantly less likely to have MACE within 730 days (HR =0.48, 95% CI = 0.43-0.53, P<0.001), as shown in Figure 4.4. However, this observation may be due to the low number of poor metabolizers with MACE events. Of the 7 poor metabolizers with MACE, 6 had MACE occur within 33 days or less (one poor metabolizer did not have MACE occur until 321 days). To further assess the impact of metabolizer status, we performed a separate analysis based on those non-responders who had MACE occur within 30 days (N = 305). In this case, high-risk poor metabolizers were significantly more likely to experience MACE within the first 30 days (HR = 3.13, 95% CI = 2.16-4.54, P < 0.001). Figure 4.5 shows a Kaplan-Meier curve of MACE within the first 30 days, stratified by metabolizer status.

Figure 4.5 Kaplan-Meier curve of non-responders within first 30 days, stratified by CYP2C19 metabolizer status



4.3 Materials & Methods

4.3.1 Study Population

Subjects in this study represent a subset of the RPGEH GERA cohort. A detailed description of the GERA cohort has been previously published[15] and is included in the Materials & Methods sections of the previous chapters, as well as at dbGAP, accession number phs000674.v1.p1. At the time of enrollment, GERA members were adult members of Kaiser Permanente Medical Care Plan, Northern California Region with high-density genotype data linked to Kaiser Permanente EHR. For this chapter, only GERA members with evidence of MI or PCI, as well as at least one outpatient clopidogrel prescription within 30 days of the MI or PCI (as described in 4.3.2 Phenotype) were included in the cohort for this study.

4.3.2 Phenotype

We first identified all GERA members with at least one outpatient clopidogrel prescription. Cloidogrel records were extracted from dispensed outpatient pharmacy records for all drugs with an active ingredient matching "clopidogrel" in the generic label name, and drug class matching "platelet aggregation inhibitors". Records were filtered to only include those dispensed and not returned to stock. Prescription records for clopidogrel were available starting in 1997, and were queried up to and including January 1, 2012.

We obtained records of PCI procedures with a drug-eluting stent for all GERA members, as well as any records of MI diagnosis. An ICD-9 code of "92980" or "92981", or CPT code of "C1874", "C1875", "C1876", or "C1977" was used to define a PCI procedure. An ICD-9 code of "410", which includes all subcodes (e.g. "410.1"), was used to identify MI events. Any PCI or MI event records from the emergency department were excluded, as we wanted to capture initial events that resulted in hospitalizations and verified as principal diagnoses.

Only GERA members with an MI, PCI, or both, as well as evidence of at least one clopidogrel prescription within 30 days of the initial event were kept for further analysis .To identify non-responders, we obtained all MACE records for the cohort. All events must have taken place within 730 days of the first clopidogrel prescription following the initial event. All events records from the emergency department were excluded. MACE ICD-9 and CPT codes observed in the GERA cohort are listed in Table 4.2 and include codes for MI, coronary artery bypass grafting, angioplasty, stroke, and a range of other cardiac-related diagnoses. For a cleaner response phenotype, we excluded from analysis individuals with evidence of MACE after 730 days following the first clopiodgrel prescription. A schematic of the phenotype is shown in Figure 4.5.

A multivariate Cox regression analysis of time to MACE was performed with a time-dependent covariate of cumulative clopidogrel exposure and Efron approximation, adjusting for CYP2C19 metabolizer status, ethnicity, age at initial event, smoking status, sex, and type of initial event. Exposure and time to MACE were calculated for up to 730 following the first clopidogrel prescription after the initial event. We calculated days to MACE starting from the date of the first clopidogrel prescription following the initial MI or PCI event. Cumulative clopidogrel exposure in days was calculated beginning at the first clopidogrel prescription after the initial MI or PCI event, and ending at either a MACE event or 730 days for responders. CYP2C19 metabolizer status was assigned based on CYP2C19 star allele diplotypes derived in Chapter 2, according to CPIC guidelines as shown in Table 4.3. Data cleaning, processing, phenotype extraction, and all analyses were performed in R version 3.0.2[16].

Table 4.2 Diagnosis and procedure codes used to define MACE events

MACE	ICD-9 Code	CPT Code
Myocardial Infarction	410*	--
Coronary Artery Bypass Grafting	--	33533, 33522, 33521, 33519, 33517, 33534, 33518, 33523, 33510, 33536, 33535, 33513, 33511, 33514, 33516, 33512
Angioplasty	--	92982, 92984, 92995, 92996
Stroke	436, 437.9, 435.9, 434.90, 437.7, 433.10, 435.0, 430, 437.1, 437.3, 433.30, 433.11, 434.91, 434.01, 437.4, 433.90, 437.0, 431, 437.8, 432.9, 434.11, 432.1, 433.80, 433.20, 435.2, 435.3	--
Other (includes Unstable Angina, Congestive Heart Failure, Malignant Dysrhythmia)	428.0, 427.9, 427.31, 427.1, 426.7, 426.11, 426.2, 427.32, 426.51, 427.69, 427.89, 427.61, 427.41, 427.0, 427.81, 428.1, 426.4, 427.5, 426.3, 426.0, 428.9, 428.20, 428.42, 428.30, 426.53, 426.10, 426.52, 427.60, 428.22,	

	428.33, 428.32, 428.21, 428.23, 428.43, 426.82, 427.42, 428.41, 426.6, 428.31, 426.9, 426.13, 427.2, 428.40, 426.12, 426.54, 428, 411.1, 411.89, 411, 411.81	

Figure 4.6 Phenotype schematic

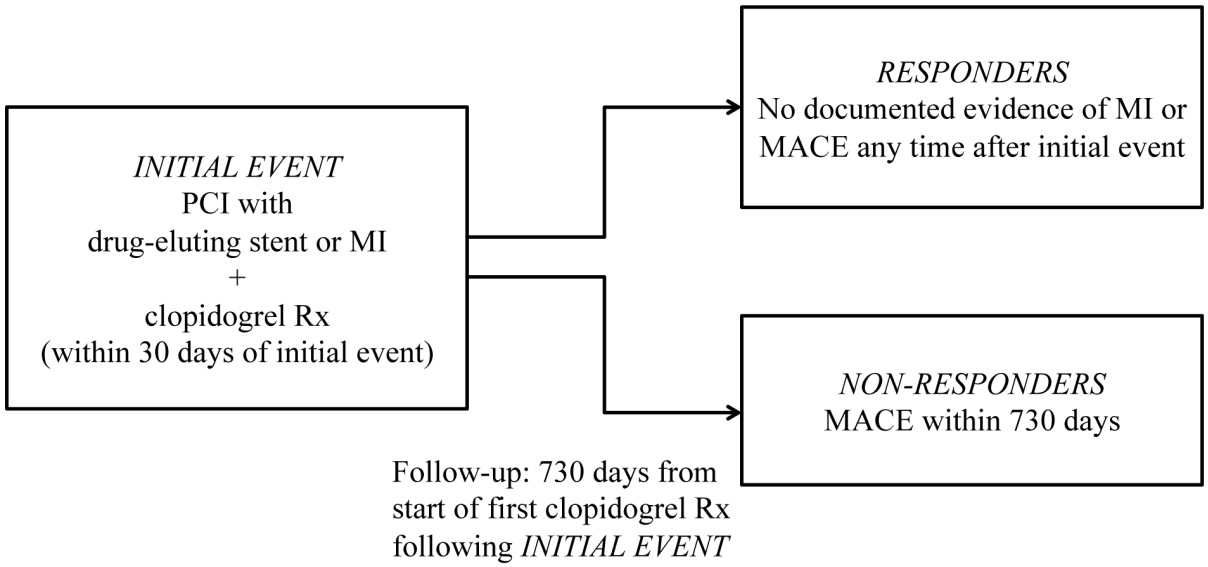


Table 4.3 Clopidogrel CYP2C19 metabolizer status by star allele diplotype

Star Alelle Diplotype	Metabolizer Phenotype	Recommended Action	Metabolizer Status Phenotype
*1/*1	EM	Standard dosing of clopidogrel[17]	Standard Therapy
*1/*17	UM		
*17/*17	UM		
*1/*2	IM	Consider alternative antiplatelet agent (e.g. prasugrel, ticagrelor) [17]	Actionable
*1/*3	IM		
*2/*17	IM		
*2/*2	PM		Actionable (High-Risk)
*2/*3	PM		
*3/*3	PM		

Metabolizer status phenotypes defined by CPIC guidelines and star allele diplotypes. EM = extensive metabolizer, UM = ultra-rapid metabolizer, IM = intermediate metabolizer, PM = poor metabolizer. Star allele translation is discussed in greater detail in Chapter 2.

4.3.3 Genotype Data

This chapter leverages genotype information previously generated for the Pharmacogene Characterization study presented in Chapter 2. Genotyping methods are described in 2.3 Materials & Methods. Briefly, GERA cohort members were genotyped on four custom Affymetrix Axiom arrays created to capture genome-wide variation in the four major race-ethnicity groups in the cohort: African Americans, East Asians, Latinos, and Non-Hispanic Whites[18,19]. Arrays included 674-893k SNPs, depending on the array[18,19]. Ethnicity assignment was largely based on self-report. Using CPIC guidelines and public pharmacogenetic databases, genotype data was converted to star allele nomenclature, and diplotypes were obtained for all cohort members. *CYP2C19* star allele diplotypes for the study cohort were derived as described in Chapter 2 and are utilized in this chapter, along with corresponding metabolizer status phenotypes, as shown in Table 4.3.

4.4 Discussion

This study presents analysis results of a drug response phenotype derived solely from EHR data. Using recurrent MACE as an indirect measure for clopidogrel non-responsiveness, we observed a non-responder frequency of 25%, which is on the lower end of what has been reported in literature[2-4]. However, we did not see significant differences in MACE occurrence based on *CYP2C19* star allele diplotype in a one or two year follow-up period. While many studies have reported this as a strong genetic association, differences in phenotype definition may play a large role in determining the effect size. One meta-analysis of 15 studies of clopiodgrel response reported a lack of substantial or consistent genetic association[20].

In designing phenotypes derived from an EHR, there are several decisions to make in terms of cutoff values, follow-up periods, and inclusion and exclusion criteria that will affect the

final sample size. For example, increasing the follow-up period from 1 to 2 years increased the number of non-responders in our cohort. We also initially assessed factors such as whether an active prescription was observed within 30 days of MACE (31% of non-responders met this criterion), but found that such filters may be too stringent. We expected to observe a higher rate of cumulative clopiodgrel exposure, especially for patients who had undergone PCI for which 1 year of clopidogrel and aspirin is recommended as standard of care. However, based on outpatient prescription records, we only observed that 59% of patients with stents had over 300 days of clopidogrel. Defining phenotypes is a substantial challenge when utilizing the EHR for pharmacogenetic research.

In this chapter, we sought to capture the clinical effect of clopidogrel on platelet aggregation, while still maintaining a relatively large sample size and data that reflected real-world observations. Certain findings, such as our analysis based on metabolizer status phenotype, warrant further investigation. For example, we found that those with high-risk poor metabolizer phenotypes actually had a lower rate of MACE compared to those with standard or actionable metabolizer phenotypes, within the 2-year follow up period. While unexpected, further analysis of the first 30 days revealed that poor metabolizers were much more likely to experience MACE (HR = 3.13) shortly after discharge. While this association was significant, there were only 7 poor metabolizer individuals who experienced MACE (only 3% of the total cohort are poor metabolizers, or 1.7% of all non-responders), and 6 out of those 7 experienced MACE within 33 days. We still did not observe a significant difference between actionable and standard metabolizers within this 30-day period, or within the total 730-day follow-up period. At a larger scale, these findings suggest that metabolizer status phenotype may be an important

factor to consider when trying to predict or reduce 30-day hospital readmission rates in this patient population.

While this chapter provides a foundation for future analyses of this phenotype, several clinical covariates were not included and may be important to consider. These include estimated glomular filtration rate and systolic blood pressure, comorbidities like hypertension, diabetes, or hyperlipidemias, as well as concomitant medications like statins, beta-blockers, and proton pump inhibitors, some of which are known to interact with *CYP2C19*. While the genetic information in this study was restricted to *CYP2C19*, the only gene with clinical pharmacogenetic guidelines available for clopidogrel, further analysis may include a genome-wide assessment of other genetic variants. As reported in Chapter 2, Asians have a larger proportion of poor metabolizers compared to other ethnic groups, but only represented 6% of this cohort. Further study of clopidogrel response in this ethnic group, perhaps in a larger replication cohort, may yield results of clinical relevance.

Although CPIC guidelines suggest alternate antiplatelet therapies for those with actionable or high-risk metabolizer phenotypes, we observed less than 10 prescriptions for prasugrel or ticagrelor. While these numbers may have increased since 2012, the data available suggests that genetic information may be of limited utility for this cohort. Further analysis is necessary to determine whether these alternative therapies, which are more costly and may have higher risks of bleeding events, are ultimately worth the cost of genotyping and pharmacogenetic integration into the EHR. The observed prescription rates may be due to institutional prescribing guidelines favoring use of clopidogrel. Additional research into the costs associated with MACE, or generally, any adverse drug event associated with drug response, is also necessary to understand the impact of genetic information on the healthcare system as a whole.

4.5 References

1. O'Riordan, M. So long, Plavix, what a ride! Clopidogrel patent expires. http://www.theheart.org/article/1400065.do (2012).
2. Mega, J., Close, S., Wiviott, S. & Shen, L. Cytochrome P-450 Polymorphisms and Response to Clopidogrel. *N Engl J Med* **360,** 354–362 (2009).
3. Snoep, J. D. *et al.* Clopidogrel Nonresponsiveness In Patients Undergoing Percutaneous Coronary Intervention With Stenting: A Systematic Review And Meta-Analysis. *Am Heart J* **154,** 221–231 (2007).
4. Scott, S. A. *et al.* Clinical Pharmacogenetics Implementation Consortium Guidelines for Cytochrome P450-2C19 (CYP2C19) Genotype and Clopidogrel Therapy. *Clin Pharmacol Ther* **90,** 328–332 (2011).
5. Momary, K. M., Dorsch, M. P. & Bates, E. R. Genetic Causes Of Clopidogrel Nonresponsiveness: Which Ones Really Count? *Pharmacotherapy* **30,** 265–274 (2012).
6. Fefer, P. & Matetzky, S. The Genetic Basis Of Platelet Responsiveness To Clopidogrel. A Critical Review Of The Literature. *Thromb Haemost* **106,** 203–210 (2011).
7. Beitelshees, A. L., Horenstein, R. B., Vesely, M. R., Mehra, M. R. & Shuldiner, A. R. Pharmacogenetics And Clopidogrel Response In Patients Undergoing Percutaneous Coronary Interventions. *Clin Pharmacol Ther* **89,** 455–459 (2011).
8. Angiolillo, D. J. *et al.* Lack Of Association Between The P2Y12 Receptor Gene Polymorphism And Platelet Response To Clopidogrel In Patients With Coronary Artery Disease. *Thromb Res Suppl* **116,** 491–497 (2005).
9. Smith, S. M. *et al.* Common Sequence Variations In The P2Y12 And CYP3A5 Genes Do Not Explain The Variability In The Inhibitory Effects Of Clopidogrel Therapy. *Platelets*

17, 250–258 (2006).

10. Lev, E. I. *et al.* Genetic Polymorphisms Of The Platelet Receptors P2Y12, P2Y1 And GP Iiia And Response To Aspirin And Clopidogrel. *Thromb Res* **119,** 355-360 (2007).
11. Simon, T. *et al.* Genetic Determinants of Response to Clopidogrel and Cardiovascular Events. *N Engl J Med* **360,** 363–375 (2009).
12. Collet, J.-P. *et al.* Cytochrome P450 2C19 Polymorphism In Young Patients Treated With Clopidogrel After Myocardial Infarction: A Cohort Study. *Lancet* **373,** 309–317 (2009).
13. Shuldiner, A. R. *et al.* Association of Cytochrome P450 2C19 Genotype With the Antiplatelet Effect and Clinical Efficacy of Clopidogrel Therapy *JAMA* **302,** 849–857 (2009).
14. U.S. Food and Drug Administration. FDA Drug Safety Communication: Reduced Effectiveness Of Plavix (Clopidogrel) In Patients Who Are Poor Metabolizers Of The Drug. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm203888.htm (2010).
15. Kvale, M. N. *et al.* Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200,** 1051–1060 (2015).
16. R Core Team. R: A Language And Environment For Statistical Computing. R Foundation For Statistical Computing. http://www.r-project.org/ (2015).
17. Scott, S. A. *et al.* Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C19 Genotype and Clopidogrel Therapy: 2013 Update. *Clin Pharmacol Ther* **94,** 317–323 (2013).

18. Hoffmann, T. J. *et al.* Design And Coverage Of High Throughput Genotyping Arrays Optimized For Individuals Of East Asian, African American, And Latino Race/Ethnicity Using Imputation And A Novel Hybrid SNP Selection Algorithm. *Genomics* **98,** 422–430 (2011).

19. Hoffmann, T. J. *et al.* Next Generation Genome-Wide Association Tool: Design And Coverage Of A High-Throughput European-Optimized SNP Array. *Genomics* **98,** 79–89 (2011).

20. Bauer, T. *et al.* Impact Of CYP2C19 Variant Genotypes On Clinical Efficacy Of Antiplatelet Treatment With Clopidogrel: Systematic Review And Meta-Analysis. *BMJ* **343,** d4588–d4588 (2011).

5 CONCLUSION

This work focused on the potential impact of combining genetic information with electronic health records (EHRs). The rapidly decreasing costs of genotyping and sequencing technology, combined with the increasing rate of adoption of EHR systems, allow for an unprecedented opportunity to assess the impact of genetics in a clinical context. This work, based on a cohort of over 100,000 individuals, provides a foundation for larger precision medicine studies. In this dissertation, we show the utility of genetic information in the context of drug prescribing. Even for a small number of drugs, this information can potentially affect thousands of individuals in our cohort. If we broaden our scope to include disease risk and more complex phenotypes, the impact of genetic information in the EHR will likely magnify.

As addressed in this dissertation, there are multiple challenges in integrating genetic data into the EHR, at multiple levels. For one, data collection is a non-trivial task that must meet several regulatory requirements, especially if the data are to be used clinically. Once collected, several informatics challenges exist, including storage of the data in a secure data warehouse, data standards, and ability to connect to an institution's EHR system[1]. As addressed in Chapter 2, nomenclature differences require conversion of genetic data into a clinically meaningful format. Beyond this, guidelines and clinical decision support (CDS) tools are necessary to translate this information into knowledge or recommendations. Such guidelines may vary by institution and should be continuously reviewed as pharmacogenetic evidence grows[2]. This is especially important as the field shifts towards next-generation sequencing technologies that may identify a large number of genetic variants of unknown clinical significance in patients.

At a broader level, physician education and medical school training regarding the role of genetics in medicine is a necessary step for wide-scale adoption of pharmacogenetics[3-5].

Additionally, reimbursement presents a major challenge for both payers and providers. If a healthcare system experiences high patient churn, the upfront costs of capturing genetic data may not be worth the implementation. Additionally, once data is generated, questions of ownership, privacy, and security may arise, especially as healthcare systems move towards improving patients' access to their own clinical data.

Despite these hurdles, integrating genetic information into the EHR has the potential to greatly impact clinical practice at point of care, particularly in the case of drug prescribing. Traditional implementation of pharmacogenetics involves ordering a lab test, assaying individual or panels of variants at a time, then waiting for results to be returned. Such practice can be costly and inefficient[6,7] and requires DNA to be collected near the time of ordering. Returned results may not be in a format amenable to long-term storage, and thus may not be incorporated into the EHR for future use[1]. With EHR systems linked to biorepositories that have generated genotypic information, it may be possible to preemptively genotype patients for known drug-gene pairs before these drugs are administered[8,9]. Theoretically, preemptive genotyping for many well-studied drugs makes it possible to predict how patients will respond before they are prescribed a drug[10]. Then, upon making a treatment decision, a physician can reference a patient's pharmacogenetic information in the context of their current clinical state, at point-of-care. Using pharmacogenetic information and CDS, physicians may then consider alternate therapies or dosing for non-responders, which can help to avoid "trial-and-error" therapy, adverse outcomes, and hospitalizations with high healthcare costs.

As discussed in Chapters 3 and 4, EHRs can be powerful tools for extracting clinically observed phenotypes for pharmacogenetic research. However, while assessing two drug response phenotypes, we encountered several challenges. One of the largest challenges was determining

the appropriate phenotype definition. For certain phenotypes, such as disease state or physiological properties that are well-captured by lab values, defining phenotypes from the EHR can be relatively straightforward. For other phenotypes like drug response or time-dependent complex phenotypes, determining a phenotype from EHR data can be more difficult. For drug response phenotypes, clinical trials and literature were the most useful starting points for putting together a phenotype definition. However, clinical trials have rigorously defined data points to collect, which may not be available in EHR data, since the latter is largely made up of administrative ICD-9-based billing codes for diagnoses and procedures[11]. Additionally, EHR data can be sparse, especially for younger, healthy individuals who do not have many doctor visits. The data available may not be a direct reflection of the biology we wish to capture; instead, signs and symptoms that present clinically may be used as an indirect representation of the actual phenotype, as was done in Chapter 4 for clopidogrel response (rather than the more direct measure of *in vitro* platelet aggregation). In designing phenotypes based on EHR data, we found that it is crucial to review collected data and proposed phenotyping algorithms with physician experts for ongoing feedback and improvement.

While an EHR system can be a rich resource for collecting clinically observed phenotypes, we found that there are certain factors that are difficult to capture, especially in regards to drug response. Medication compliance is a major issue, especially for drugs like the atypical antipsychotics (AAPs) discussed in Chapter 3. While we are able to tell that a drug prescription is picked up, there is no record of whether the patients are actually taking the medication correctly. Additionally, for certain drugs or drug classes, it is difficult to capture exactly why a drug was prescribed. For example, we found that for certain AAPs, the prescribed dosage was much lower than expected for patients with a psychiatric condition. Upon consulting

with psychiatrists, we found that many were prescribed a lower dose for insomnia, an off-label use of the drug that was not considered in the original phenotyping algorithm.

Given these challenges, replicating phenotypes in other institutions with EHRs may be a difficult task, but would help facilitate a broader assessment of diverse patient populations and increase power for studying phenotypes such as drug response. Resources for sharing phenotyping algorithms have been set up[12], highlighting the increasing use of EHR systems for deriving phenotypes for research use. Data standards and vocabularies for describing clinical data are key to such algorithms. For prospective collection of data for phenotyping purposes, standards, vocabularies, and ontologies need to be better incorporated into EHR systems to map incoming patient data to relevant phenotypic concepts. Natural language processing tools and machine learning algorithms can be used to better capture clinical context, improve sensitivity and specificity of phenotypes, and identify distinct patient groups. On the front end, EHR systems need to be minimally disruptive, providing tailored guidance and alerts only when they are most likely to be utilized, with user interfaces that facilitate capture of information.

Despite these challenges, we were able to utilize EHR data to extract meaningful phenotypes that reflected rates shown in literature for two drug response phenotypes. Using the genetic data available for the GERA cohort, we were able to quantify variability among ethnicities in key pharmacogenes, which largely reflected reported frequencies for well-studied populations with frequency information available. By combining these resources, we found that the majority of the GERA cohort has genetic variants that may impact drug prescribing, highlighting the potential utility of genetic information in clinical practice.

5.1 References

1. Potamias, G. *et al.* Deciphering Next-Generation Pharmacogenomics: An Information

Technology Perspective. *Open Biol* **4,** 140071 (2014).

2. Crews, K. R., Hicks, J. K., Pui, C. H., Relling, M. V. & Evans, W. E. Pharmacogenomics and Individualized Medicine: Translating Science Into Practice. *Clin Pharmacol Ther* (2012).
3. Crews, K. R. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines For Codeine Therapy In The Context Of Cytochrome P450 2D6 (CYP2D6) Genotype. *Clin Pharmacol Ther* **91,** 321–326 (2012).
4. Baudhuin, L. M., Langman, L. J. & O'Kane, D. J. Translation of Pharmacogenetics into Clinically Relevant Testing Modalities. *Clin Pharmacol Ther* **82,** 373–376 (2007).
5. Scott, S. A. Personalizing Medicine With Clinical Pharmacogenetics. *Genet Med* **13,** 987–995 (2011).
6. Pulley, J. M. J. *et al.* Operational Implementation Of Prospective Genotyping For Personalized Medicine: The Design Of The Vanderbilt PREDICT Project. *Clin Pharmacol Ther* **92,** 87–95 (2012).
7. Wagner, M. J. Pharmacogenetics And Personal Genomes. *Per Med* **6,** 643–652 (2009).
8. Bielinski, S. J. *et al.* Preemptive Genotyping For Personalized Medicine: Design Of The Right Drug, Right Dose, Right Time-Using Genomic Data To Individualize Treatment Protocol. *Mayo Clin Proc* **89,** 25–33 (2014).
9. Johnson, J. A. *et al.* Implementing Personalized Medicine: Development of a Cost-Effective Customized Pharmacogenetics Genotyping Array. *Clin Pharmacol Ther* **92,** 437–439 (2012).
10. Bell, G. C. *et al.* Development And Use Of Active Clinical Decision Support For Preemptive Pharmacogenomics. *J Am Med Inform Assoc* **21**, e93-99 (2013).

11. Jensen, P. B., Jensen, L. J. & Brunak, S. Mining Electronic Health Records: Towards Better Research Applications And Clinical Care. *Nat Rev Genet* **13,** 395–405 (2012).

12. Vanderbilt University. PheKB: A Knowledgebase for Discovering Phenotypes from Electronic Medical Records. https://phekb.org (2014).

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature  Date 9/8/15