THE CHALLENGES IN USING ELECTRONIC HEALTH RECORDS FOR PHARMACOGENOMICS AND PRECISION MEDICINE RESEARCH

SARAH M. LAPER

Eastern Virginia Medical School, Norfolk, VA, 23507, USA
Email: lapersm@evms.edu

NICOLE A. RESTREPO

Center for Human Genetics Research, Vanderbilt University, 519 Light Hall, 2215 Garland Avenue, Nashville, TN 37232, USA
Email: n.restrepo@vanderbilt.edu

DANA C. CRAWFORD

Institute for Computational Biology, Department of Epidemiology and Biostatistics, Case Western Reserve University, Wolstein Research Building, 2103 Cornell Road, Suite 2527, Cleveland, OH 44106, USA
Email: dana.crawford@case.edu

Access and utilization of electronic health records with extensive medication lists and genetic profiles is rapidly advancing discoveries in pharmacogenomics. In this study, we analyzed ~116,000 variants on the Illumina Metabochip for response to antihypertensive and lipid lowering medications in African American adults from BioVU, the Vanderbilt University Medical Center's biorepository linked to de-identified electronic health records. Our study population included individuals who were prescribed an antihypertensive or lipid lowering medication, and who had both pre- and post-medication blood pressure or low-density lipoprotein cholesterol (LDL-C) measurements, respectively. Among those with pre- and post-medication systolic and diastolic blood pressure measurements (n=2,268), the average change in systolic and diastolic blood pressure was -0.6 mg Hg and -0.8 mm Hg, respectively. Among those with pre- and post-medication LDL-C measurements (n=1,244), the average change in LDL-C was -26.3 mg/dL. SNPs were tested for an association with change and percent change in blood pressure or blood levels of LDL-C. After adjustment for multiple testing, we did not observe any significant associations, and we were not able to replicate previously reported associations, such as in *APOE* and *LPA*, from the literature. The present study illustrates the benefits and challenges with using electronic health records linked to biorepositories for pharmacogenomic studies.

1. Introduction

As the costs for genomic sequencing continue to decrease, the goals of precision medicine to use a patient's genetic makeup to calculate his or her risk of disease, prevent illness, and determine the best drug or treatment for his or her medical problem, become more feasible [1,2]. Large-scale genome-wide association studies (GWAS) and smaller sequencing studies are rapidly identifying common and rare genetic variants that contribute to human disease [3] and response to drug treatment [4]. The results from pharmacogenomic studies, the study of variants that effect drug response and efficacy, can be translated to clinical practice more easily than variants that effect

disease risk. For example, knowledge that one treatment is most effective for one genotype group, while another treatment is optimal for another group, can aid in the selection of treatments [5]. There have already been successful pharmacogenomic studies that have been translated to clinical practice. A variant in *HLA-B* was identified that is associated with increased risk of a hypersensitivity reaction when using Abcavir for the treatment of HIV [6], dosing recommendations for thiopurines have been developed based on *TPMT* genotype [7], and variants in *CYP2D6* have been identified that cause patients to either be poor metabolizers or rapid metabolizers of codeine [8]. Many of the early pharmacogenomic studies focused on variants in candidate genes that code for drug-metabolizing enzymes or drug targets. However, with advances in molecular assaying technology and the increased practicality of sequencing the entire genome, variants in other regions that have a clinically important effect may be discovered [9].

The majority of genetic association studies, including pharmacogenomic studies [10,11], have been in European populations [12]. It is important to conduct GWAS in diverse populations in order to discover variants that may not be present in European populations [12]. Previous studies have already found population specific frequencies for variants that effect drug response. For example, it has been found that there are significant differences in allele frequencies between populations for genes encoding drug metabolizing enzymes [13], that variants in *CYP2C9* and *VKORC1* differ among racial/ethnic groups and effect the dosing of warfarin [14], and that African Americans have the lowest frequency of the variant near the *IL28B* gene that is associated with response to hepatitis C treatment [15].

Longitudinal epidemiological cohorts are the gold standard for genetic association studies particularly in the context of gene-environment studies [16]. Properly designed cohorts, however, require enormous resources for the study of common health outcomes and may not be feasible for the study of rare outcomes, such as adverse events in pharmacogenomics. The recent emergence of electronic health records (EHR) linked to biorepositories offers an alternative strategy for rapid and cost-effective data collection for genetic association studies. EHRs contain a large amount of patient data, and it has been shown that, when linked to biorepositories, this data source can be utilized in genetic studies [17]. The use of EHRs linked to biorepositories has advantages over the traditional cohort design, such as cost, timeliness, and the ability to select for a wide range of phenotypes [18]. Also, EHRs contain data not typically collected in a traditional epidemiological study, such as information related to drug response [5]. Extracting medication from EHRs has been found to be one of the most time-consuming processes when using EHR driven genomic studies. However, advances in natural language processing have been successful in identifying medication relevant information from clinical notes in EHRs [19]. Finally, an advantage of using EHRs is that they provide a more accurate representation of the clinical population, including minority populations, than traditional cohort studies [18].

In this study, we used EHRs linked to a biorepository to analyze drug response in an African American population of almost 12,000 patients genotyped on the Illumina Metabochip [20]. We extracted data related to two common clinical treatments: 1) the use of antihypertensive medication to lower blood pressure, and 2) the use of lipid lowering medication to lower blood

levels of low-density lipoprotein cholesterol (LDL-C). Individual response to both of these treatments varies greatly, although the exact cause of this variation is unknown and likely due to many interacting factors. The availability of EHR data allowed us to study drug response in an African American population. However, this study provides an illustration of challenges that arise when using EHRs linked to biorepositories for genetic association analyses.

2. Methods

2.1. *Study population*

The data described here were obtained from BioVU, the Vanderbilt University Medical Center's biorepository linked to de-identified electronic health records. BioVU operations [21] and ethical oversight [22] have been described elsewhere. Briefly, DNA is collected from discarded blood samples remaining after routine clinical testing at Vanderbilt outpatient clinics in Nashville, Tennessee and surrounding areas, and is linked to a de-identified version of the patient's EHR termed the "Synthetic Derivative." The data were de-identified in accordance with provisions of Title 45, Code of Federal Regulations, part 46 (45 CFT 46), and this study was considered non-human subjects research by the Vanderbilt University Internal Review Board.

2.2. *Genotyping*

DNA samples from mostly non-European Americans in BioVU were genotyped on the Illumina Metabochip by the Vanderbilt University Center for Human Genetics Research DNA Resources Core as part of the Population Architecture using Genomics and Epidemiology (PAGE) I Study [23]. This dataset is herein after referred to as "EAGLE BioVU" [24]. The Metabochip is a custom array of approximately 200,000 SNPs designed for replication and fine-mapping of genome-wide association study-identified variants for metabolic and cardiovascular traits [25].

Prior to analyses, quality control procedures were performed using PLINKv1.09 [26]. SNPs were filtered to exclude those with low minor allele frequency (<1%), low genotyping frequency (<95%), and deviations from Hardy-Weinberg expectations ($p<10^{-6}$). After quality control, approximately 116,000 SNPs were available for analysis. Patients were excluded from analysis if his or her biological sex did not match the recorded sex in the EHR, if they had a low genotyping rate (<95%), or if they were determined to be related to another sample (identical-by-descent).

2.3. *Phenotyping*

BioVU de-identified EHRs contain structured (such as International Classification Disease codes or billing codes, current procedural terminology codes, vital signs, and labs) and unstructured (clinical notes) data accessible for electronic phenotyping. We extracted systolic blood pressure, diastolic blood pressure, and LDL-C from the de-identified EHRs for the present study. Prescription medication is available in the de-identified version of the EHR through MedEx [19]. Clinic dates and corresponding prescriptions for both antihypertensives and LDL-C lowering

medications were extracted as previously described [24,27,28]. Blood pressure and LDL-C measurements were considered "post-medication" if a prescription for an antihypertensive or lipid lowering drug was extracted prior to the lab. "Pre-medication" values include all values available per patient prior to a prescription, including values within the normal range. Median values of measurements were calculated for both "pre-medication" and "post-medication" categories for each patient. Body mass index was extracted from the EHR and cleaned as previously described [29]. This study only considered patients who had both pre-and post-medication median measurements for systolic and diastolic blood pressure, or LDL-C.

EAGLE BioVU has a total of 15,863 patients [24]. For the present study, only patients who were identified as African American and who were over the age of 18 by the year 2010, based on date of birth, were included. There were 2,653 and 1,244 patients who had both pre- and post-medication systolic and diastolic blood pressure measurements and pre- and post-medication LDL-C lab values, respectively. After quality control, there were 2,268 and 1,028 patients analyzed for blood pressure medication response and lipid medication response, respectively.

2.4. *Statistical Methods*

Single SNP tests of association assuming an additive genetic model using linear regression were performed using PLINKv1.09 [26]. Two dependent variables were considered: 1) the difference between the median post- and pre-medication measurements and 2) the percent change between the median post- and pre-medication measurements. For each dependent variable, three different models were considered: 1) unadjusted, 2) adjusted for age and sex and 3) adjusted for age, sex, and the first three principal components of ancestry. Age was the patient's age as of 2010. The principal components of ancestry were obtained from the PAGE I Study Coordinating Center [30]. Results from tests of association for *APOE* and *LPA* gene regions were visualized using LocusZoom [31].

3. Results

3.1. *Response to Anti-Hypertensive Medication*

Study population characteristics for patients with pre- and post-antihypertensive medication blood pressure measurements are given in Table 1. A total of 2,620 adult African American patients had both a pre- and post-antihypertensive medication blood pressure measurement. The majority of the patients were female (65.5%) and born between 1940 and 1970 (70.9%). On average, the patients were overweight (BMI = 29.7 kg/m^2). For systolic blood pressure, the average median pre-medication measurement was 132.6 mm Hg, reflecting the fact that all pre-medication values over the course of a patient's care including those within the normal range were included. The average median post-medication measurement was 132.0 mm Hg, and the average change in systolic blood pressure with the use of medication was a decrease by 0.6 mm Hg. For diastolic blood pressure, the average median pre-medication measurement was 79.7 mm Hg, the average median post-

medication measurement was 79.0 mm Hg, and the average change in diastolic blood pressure with the use of medication was a decrease by 0.8 mm Hg.

Table 1. Blood Pressure Medication Response Study Population Characteristics. Study population characteristics (sex, decade of birth, average median body mass index, average median pre-medication blood pressure measurement, average median post-medication blood pressure measurement, and average change in blood pressure are given for the 2,620 patients who were above the age of 18, African American, and who had both pre-medication and post-medication blood pressure measurements. Abbreviations: standard deviation (SD).

Female, %	65.5
Decade of birth, %	
1910	0.7
1920	3.1
1930	8.4
1940	16.1
1950	24.6
1960	20.5
1970	13.3
1980	12.1
1990	1.2
Mean (± SD) Body Mass Index (kg/ m^2)	29.2 (12.7)
Mean (± SD) pre-medication blood pressure measurement (systolic/ diastolic) (mm Hg)	132.6 (17.1) / 79.7 (10.6)
Mean (± SD) post-medication blood pressure measurement (systolic/diastolic) (mm Hg)	132.0 (15.0) / 79.0 (9.4)
Mean (± SD) change in blood pressure (systolic/ diastolic) (mm Hg)	-0.6 (15.2) / -0.8 (10.0)

The 116,000 SNPs on the Metabochip that passed quality control, as described in the methods, were tested for an association with change and percent change in systolic and diastolic blood pressure with the use of antihypertensive medication. No SNPs passed a strict Bonferroni corrected significance level ($p = 4 \times 10^{-7}$) for an association with either the change or percent change in systolic or diastolic blood pressure for any of the models (Figure 1). The most significant SNP for both the change and percent change in systolic blood pressure was rs8058830 on chromosome 16 (beta = -2.6 mm Hg, $p = 1.4 \times 10^{-5}$; beta = -1.9%, $p = 1.3 \times 10^{-5}$). The most significant SNP for the change in diastolic blood pressure mapped to rs183551129 on chromosome 15 (beta = 1.5 mm Hg, $p = 1.2 \times 10^{-5}$). The most significant SNP for the percent change in diastolic blood pressure was rs17672219 on chromosome 16 (beta = -2.2%, $p = 6.5 \times 10^{-}$

^{6}). These results are for the linear regression model adjusted for age, sex, and the first three principal components, and the results were not appreciably different for the other two models.

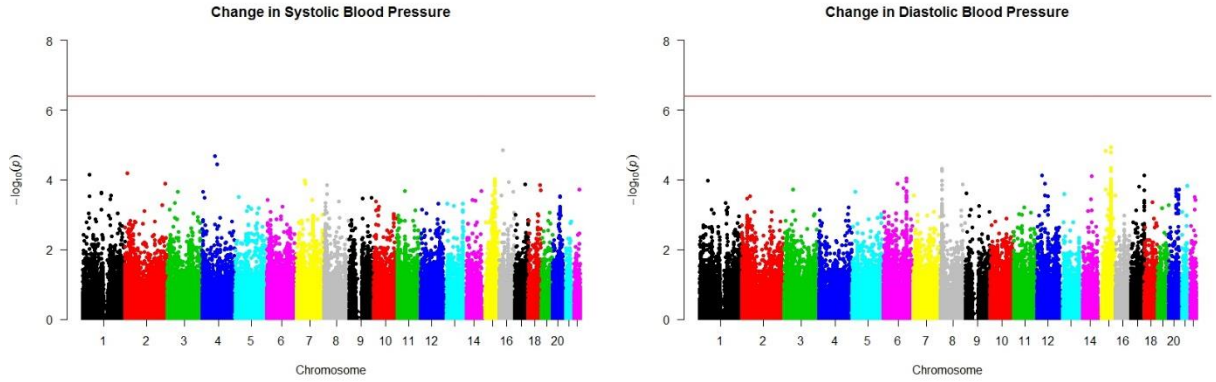

Figure 1. Single SNP association results for change in blood pressure with the use of antihypertensive medication. Single SNP tests of association were performed using linear regression assuming an additive genetic model adjusting for age, sex, and the first three principal components. The $\log_{10}(p)$ values (y-axis) were plotted using R for the association of each tested SNP (x-axis) with the change in systolic and diastolic blood pressure respectively. The red line represents the Bonferroni corrected significance level of $p = 4 \times 10^{-7}$.

3.2. *Response to Lipid Lowering Medication*

Study population characteristics for patients with pre- and post-medication LDL-C measurements are given in Table 2. A total of 1,244 patients African American adult patients had both a pre- and post-medication LDL-C lab measurement. The majority of the patients were female (61.4%) and born between 1930 and 1970 (85.8%). On average, the patients were obese (BMI = 30.9 kg/m^2). The average pre-medication LDL-C measurement was 129.9 mg/dL, the average post-medication LDL-C measurement was 103.6 mg/dL, and the average change in LDL-C with the use of lipid lowering medication was a decrease by 26.3 mg/dL.

The 116,000 SNPs on the Metabochip that passed quality control as described in the methods were tested for an association with change and percent change in LDL-C with the use of lipid lowering medication. For both the change and percent change in LDL-C, none of the SNPs passed a Bonferroni corrected significance level ($p = 4 \times 10^{-7}$) for any of the linear regression models (Figure 2). The most significant SNP for the change in LDL-C was rs12564350 located on

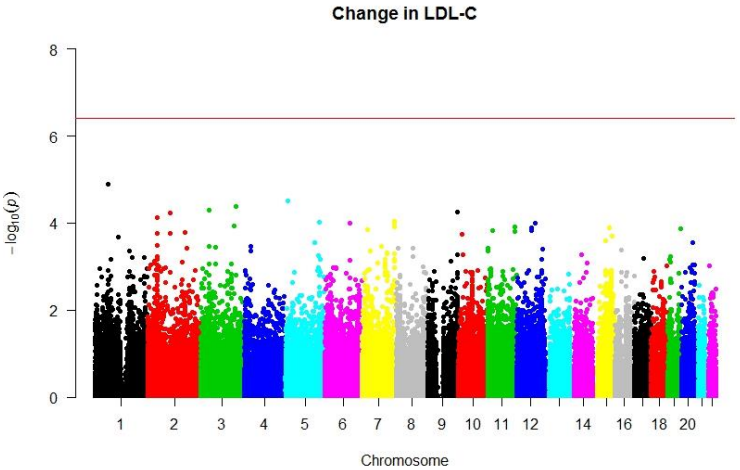

Figure 2. Single SNP association results for change in LDL-C with the use of lipid lowering medication. Single SNP tests of association were performed using linear regression assuming an additive genetic model adjusting for age, sex, and the first three principal components. The $\log_{10}(p)$ values (y-axis) were plotted using R for the association of each tested SNP (x-axis) with the change in LDL-C measurements. The red line represents the Bonferroni corrected significance level of $p = 4 \times 10^{-7}$.

chromosome 1 (beta = -14.8 mg/dL, $p = 1.3 \times 10^{-5}$). The most significant SNP for the percent change in LDL-C was rs4309741 on chromosome 3 (beta = 9.3%, $p = 5.6 \times 10^{-6}$). These results are for the linear regression model adjusted for age, sex, and the first three principal components, and the results for the other two models tested were not appreciably different.

Table 2. Statin Response Study Population Characteristics. Study population characteristics (sex, decade of birth, average median body mass index, average median pre-medication LDL-C measurement, average median post- medication LDL-C measurement, and average change in LDL-C are given for the 1,242 patients who were above the age of 18, African American, and who had both pre-medication and post-medication LDL-C measurements. Abbreviations: standard deviation (SD); low density lipoprotein cholesterol

Female, %	61.4
Decade of birth, %	
1910	1.0
1920	5.6
1930	15.7
1940	22.9
1950	27.9
1960	19.3
1970	6.0
1980	1.4
1990	0.2
Mean (± SD) Body Mass Index (kg/ m^2)	30.9 (10.9)
Mean (± SD) pre-medication blood LDL-C measurement (mg/dL)	129.9 (45.5)
Mean (± SD) post-medication blood LDL-C measurement (mg/dL)	103.6 (35.7)
Mean (± SD) change in LDL-C (mg/dL)	-26.3 (41.4)

Two genes for which variants have been located that are significantly associated with lipid response to medication and have been consistently replicated are *APOE* and *LPA* [32]. There were approximately ~250 and ~500 SNPs assayed in the *LPA* and *APOE* regions shown, respectively.

The SNPs that were tested in these regions on the Metabochip were not found to be significantly associated with the change in LDL-C with the use of lipid lowering medication (Figure 3).

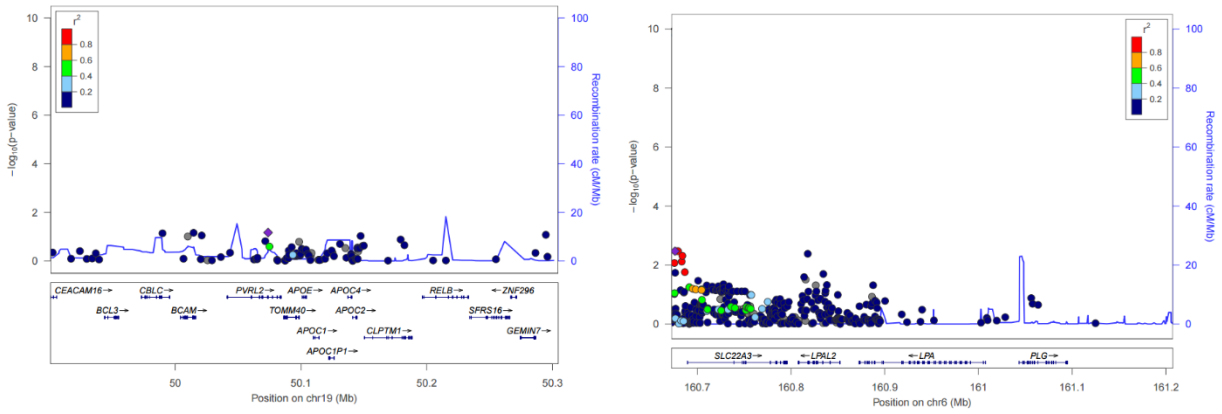

Figure 3. Locus Zoom plot of association results for *APOE* and *LPA*. Figure was generated using LocusZoom (http://csg.sph.umich.edu/locuszoom/) with linkage disequilibrium calculations from the hg18 1000 Genomes June 2010 YRI dataset. Results are shown for the association results in the *APOE* and *LPA* regions for the change in LDL-C with the use of medication, for the linear regression model adjusted for age, sex, and the first three principal components.

4. Discussion

In this study, we extracted data from BioVU related to antihypertensive and lipid lowering medication use in adult, African American patients. We tested for the association of SNPs on the Metabochip with the change in blood pressure and LDL-C measurements with the use of antihypertensive and lipid lowering medicines, respectively. Previous studies have found variants significantly associated with drug response for both of these medications [32–34]. After correction for multiple testing, we did not identify significant novel associations nor did we replicate previously reported associations for these response to treatment outcomes. There were several limitations to this study that illustrate many of the challenges for pharmacogenomics studies that use data from EHRs linked to biorepositories.

A major challenge for all pharmacogenomics studies is sample size and statistical power. The electronic phenotyping strategy outlined here balanced sample size ("lumping") versus precise phenotyping ("splitting"). In this present study, we did not distinguish by class of medication prescribed. There have been previous genome-wide association studies in African American populations that have found SNPs significantly associated with response to antihypertensive medications [33,34]. However, these studies differed from the current study in that they had much smaller sample sizes and studied specific classes of antihypertensive medication. A study on the response to thiazide diuretics had a sample size of 204 African American patients [34] and a study on angiotensin II receptor blockers used 193 African American patients [33]. Our study had more power to detect a significant association since we had 2,620 African American individuals, but

this was balanced by the fact that the Metabochip did not assay any of the specific previously identified variants and our reliance on linkage disequilibrium. Also, the strategy of lumping rather splitting medication classes may have precluded our ability to identify associations in this dataset.

There have also been previous genome-wide association studies on response to lipid lowering medications that have found significant associations [35–39]. Although these studies found significant associations for variants in several genes, only variants in the *APOE* and *LPA* genes have been consistently replicated [32]. Our study found no significant SNPs in these regions (Figure 3). However, all of the previous studies were conducted using European populations. Since our study was performed in an African American population it is not entirely surprising that we were not able to replicate previous significant associations since allele frequencies and levels of linkage disequilibrium vary between populations, risk variants vary in effect size between different populations, and there may be risk variants in African American populations that do not exist in European populations [12].

One of the limitations of this study is that genotyping was performed on the Metabochip, a custom array designed for fine mapping of variants identified by genome-wide association studies for cardiovascular and metabolic traits [25]. Although we had variants in the gene regions of significant SNPs identified in previous studies on antihypertensive and lipid lowering medication response, most of the specific previously identified SNPs were not tested. Thus, we may have missed significant associations because the SNPs were not directly genotyped or in strong linkage disequilibrium with the SNPs targeted by the Metabochip. This, indeed, is another problem in pharmacogenomic studies. Variants in regions known to effect drug metabolism often do not pass genotype quality control, and their coverage on genome-wide association study genotyping platforms is limited [40,41].

As already noted, a common problem with pharmacogenomic studies is obtaining a large enough sample size since it is difficult to obtain a large population taking a certain class of drug or having a particular response [5,40]. Even though we had access to a large number of African American patients, we still only had a small sample size for patients who met our phenotype requirements. This illustrates one of the disadvantages to using EHR data in the United States. Since it is an open system, some of the patients did not come back for a post-medication measurement or entered the system after already starting a medication.

Phenotyping is a major challenge when using EHRs for pharmacogenomics studies. We used MedEx to extract data on when a prescription for a qualifying medication was entered into the system. However, we did not distinguish between the types or doses of medication because our sample size would have been too small to have the power to detect an association. Also, EHR does not include data on many environmental factors. For example, we could not access compliance, since BioVU is an opt-out model and we could not follow-up with patients, so we assumed that if a patient had a prescription for a medication than they were actually taking the medication. One of the advantages to using EHR data is that there is access to patient information over a long time span. The biorepository we used contains patient data spanning 20 years. However, this presents a

challenge in our pharmacogenomics study because prescribing practices have changed for both blood pressure and lipid lowering medication.

As compared to a traditional epidemiological study with standard protocols aimed at uniform measurement, the quality of measurements recorded in EHRs can be variable. Blood pressure is hard to measure for numerous reasons [42]. It is possible that different instruments are used across the different outpatient clinics to measure blood pressure, and these differences can result in slightly higher or slightly lower measurements. In a clinical setting, patient's blood pressure measurements are taken at different times of the day depending on appointment time, and each person's blood pressure varies throughout the day. The white coat effect may also cause an artificially high measurement. Finally, there is inherent human error in taking blood pressure measurements. The lab values for LDL-C are more reliable, but we have to assume that the patient in BioVU fasted, and this is not typically documented in the EHR.

Although we experienced challenges in using EHR data linked to a biorepository to study drug response for antihypertensive and lipid lowering medications, this dataset did allow us to study a diverse population. The previous studies on response to antihypertensive medication had small sample sizes for African Americans and there have been no genome-wide association studies on response to lipid lowering medication in African Americans. Some of the challenges that have been described here would be difficult to overcome in our current health care system. However, simply combining EHR data from multiple locations such as EHR-linked biorepositories participating in the electronic MEdical Records & GEnomics (eMERGE) network [17] may allow us to distinguish between the type and dose of medication while maintaining adequate sample sizes. The use of EHR data is a promising and valuable resource for the future of pharmacogenomic studies in diverse populations as we enter the era of precision medicine.

5. Acknowledgements

This work was supported in part by NIH grant U01 HG004798 and its ARRA supplements. The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant funded by the National Center for Research Resources, Grant UL1 RR024975-01, which is now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06.

References

1. W. G. Feero, JAMA **299**, 1351 (2008).
2. F. S. Collins and H. Varmus, N. Engl. J. Med. **372**, 793 (2015).
3. T. A. Manolio, Nat. Rev. Genet. **14**, 549 (2013).
4. M. Pirmohamed, Annu. Rev. Genomics Hum. Genet. **15**, 349 (2014).
5. M. D. Ritchie, Hum. Genet. **131**, 1615 (2012).
6. M. A. Martin, T. E. Klein, B. J. Dong, M. Pirmohamed, D. W. Haas, and D. L. Kroetz, Clin. Pharmacol. Ther. **91**, 734 (2012).

7. M. V. Relling, E. E. Gardner, W. J. Sandborn, K. Schmiegelow, C.-H. Pui, S. W. Yee, *et al*, Clin. Pharmacol. Ther. **89**, 387 (2011).
8. K. R. Crews, A. Gaedigk, H. M. Dunnenberger, T. E. Klein, D. D. Shen, J. T. Callaghan, *et al*, Clin. Pharmacol. Ther. **91**, 321 (2012).
9. W. E. Evans and M. V. Relling, Science **286**, 487 (1999).
10. V. E. Ortega and D. A. Meyers, J. Allergy Clin. Immunol. **133**, 16 (2014).
11. L. H. Cavallari and M. A. Perera, Future Cardiol. **8**, 563 (2012).
12. N. A. Rosenberg, L. Huang, E. M. Jewett, Z. A. Szpiech, I. Jankovic, and M. Boehnke, Nat. Rev. Genet. **11**, 356 (2010).
13. J. Wilson, M. Weale, A. C. Smith, F. Gratix, B. Fletcher, M. G. Thomas, *et al*, Nature Genetics **29**, 265 (2001).
14. J. A. Johnson, L. Gong, M. Whirl-Carrillo, B. F. Gage, S. A. Scott, C. M. Stein, *et al*, Clin. Pharmacol. Ther. **90**, 625 (2011).
15. D. Ge, J. Fellay, A. J. Thompson, J. S. Simon, K. V. Shianna, T. J. Urban, *et al*, Nature **461**, 399 (2009).
16. T. A. Manolio, J. E. Bailey-Wilson, and F. S. Collins, Nat. Rev. Genet. **7**, 812 (2006).
17. D. C. Crawford, D. R. Crosslin, G. Tromp, I. J. Kullo, H. Kuivaniemi, M. G. Hayes, *et al*, Front. Genet. **5**, (2014).
18. I. S. Kohane, Nat. Rev. Genet. **12**, 417 (2011).
19. H. Xu, M. Jiang, M. Oetjens, E. A. Bowton, A. H. Ramirez, J. M. Jeff, et al, Am. Med. Inform. Assoc. JAMIA **18**, 387 (2011).
20. D. C. Crawford, R. Goodloe, E. Farber-Eger, J. Boston, S. A. Pendergrass, J. L. Haines, *et al*, Hum. Hered. **79**, 137 (2015).
21. D. Roden, J. Pulley, M. Basford, G. Bernard, E. Clayton, J. Balser, *et al*, Clin. Pharmacol. 38 Ther. **84**, 362 (2008).
22. J. Pulley, E. Clayton, G. R. Bernard, D. M. Roden, and D. R. Masys, Clin. Transl. Sci. **3**, 42 (2010).
23. T. C. Matise, J. L. Ambite, S. Buyske, C. S. Carlson, S. A. Cole, D. C. Crawford, *et al*, Am. J. Epidemiol. **174**, 849 (2011).
24. D. Crawford, R. Goodloe, E. Farber-Eger, J. Boston, S. Pendergrass, J. Haines, et al, Human Hereditary **79**, 137 (2015).
25. B. F. Voight, H. M. Kang, J. Ding, C. D. Palmer, C. Sidore, P. S. Chines, *et al*, PLoS Genet. **8**, e1002793 (2012).
26. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham, Am. J. Hum. Genet. **81**, 559 (2007).
27. L. Dumitrescu, R. Goodloe, Y. Bradford, E. Farber-Eger, J. Boston, and D. C. Crawford, BioData Min. **8**, (2015).
28. L. Dumitrescu, R. Goodloe, E. Farber-Eger, S. Pendergrass, W. Bush, and D. Crawford (submitted).
29. R. Goodloe, E. Farber-Eger, J. Boston, D. Crawford, and W. Bush, (in preparation).
30. S. Buyske, Y. Wu, C. L. Carty, I. Cheng, T. L. Assimes, L. Dumitrescu, *et al*, PLoS ONE **7**, e35651 (2012).

31. R. J. Pruim, R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, *et al*, Bioinformatics **26**, 2336 (2010).
32. J. C. Hopewell, C. Reith, and J. Armitage, Curr. Opin. Lipidol. **25**, 438 (2014).
33. S. T. Turner, E. Boerwinkle, J. R. O'Connell, K. R. Bailey, Y. Gong, A. B. Chapman, *et al*, Hypertension **62**, 391 (2013).
34. S. T. Turner, K. R. Bailey, B. L. Fridley, A. B. Chapman, G. L. Schwartz, H. S. Chai, *et al*, Hypertension **52**, 359 (2008).
35. J. F. Thompson, C. L. Hyde, L. S. Wood, S. A. Paciga, D. A. Hinds, D. R. Cox, *et al*, Circ. Cardiovasc. Genet. **2**, 173 (2009).
36. M. J. Barber, L. M. Mangravite, C. L. Hyde, D. I. Chasman, J. D. Smith, C.A. McCarty, *et al*, PLoS ONE **5**, e9763 (2010).
37. D. I. Chasman, F. Giulianini, J. MacFadyen, B. J. Barratt, F. Nyberg, and P. M. Ridker, Circ. Cardiovasc. Genet. **5**, 257 (2012).
38. H. A. Deshmukh, H. M. Colhoun, T. Johnson, P. M. McKeigue, D. J. Betteridge, P. N. Durrington, *et al*, J. Lipid Res. **53**, 1000 (2012).
39. J. C. Hopewell, S. Parish, A. Offer, E. Link, R. Clarke, M. Lathrop, *et al*, Eur. Heart J. **34**, 982 (2013).
40. A. A. Motsinger-Reif, E. Jorgenson, M. V. Relling, D. L. Kroetz, R. Weinshilboum, N. J. Cox, *et al*, Pharmacogenet. Genomics **23**, 383 (2013).
41. M. T. Oetjens, J. C. Denny, M. D. Ritchie, N. B. Gillani, D. M. Richardson, N. A. Restrepo, *et al*, Pharmacogenomics **14**, 735 (2013).
42. D. W. Jones, L. J. Appel, S. G. Sheps, E. J. Roccella, and C. Lenfant, JAMA **289**, 1027 (2003).