Residual-based Test of Conditional Association between Continuous and Ordinal Variables with Application to Genome-wide Association Studies

By

Valentine Adhiambo Wanga

Thesis

submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biostatistics

August, 2014

Nashville, Tennessee

Approved:

Bryan E. Shepherd, Ph.D.

Chun Li, Ph.D.

ABSTRACT

The discovery of genes linked with a large array of diseases has been accelerated by genome-wide association studies (GWAS), in which genetic variants in different individuals are examined for relationship with a specified phenotype. Most GWAS analyses require modeling the association between single nucleotide polymorphisms (SNPs) and the outcome of interest as additive, dominant, or recessive. In general, this relationship is not known. The genotypes of a marker can be regarded as ordered categorical. An additive model assumes linearity, and approaches that categorize the data ignore order information, resulting in loss of power. Therefore, a method that only assumes a monotonic relationship between SNPs and the outcome of interest may be more robust and powerful than standard approaches. In this thesis, we explore the use of such a method using pharmacogenomics data from a clinical trial that randomized 1858 HIV-infected patients to one of four antiretroviral regimen combinations (tenofovir+efavirenz, tenofovir+atazanavir, abacavir+efaverinz, and abacavir+atazanavir). We are specifically interested in detecting SNPs that are associated with tenofovir clearance and creatinine clearance. We assess the performance of the new method versus the additive, dominant, and recessive models via simulation studies and real data analyses, and compare and contrast findings.

ACKNOWLEGDEMENTS

I would like to thank my advisors, Dr. Bryan Shepherd and Dr. Chun Li, for their guidance and support over the years of my graduate studies at Vanderbilt. I have learned so much from them through their comments and feedback, especially on how to ask research questions even when they are not so obvious. During my work, I also had the opportunity to work with Dr. Haas, whose feedback was very helpful in understanding useful aspects in the analysis of genetic data, particularly data on HIV/AIDS patients randomized to anti-retroviral therapy. I would like to thank the Aids Clinical Trials Group (ACTG) and the participants of Trial A5202 for allowing me to use their data for my case studies.

I have been fortunate to have Dr. Jeffrey Blume as my academic advisor during my studies here at Vanderbilt. He not only believed in me, but also provided timely advice and encouragement when I needed them most.

Many thanks to Charles Dupont for his patience and the time he took to help with some software implementations, especially at the beginning of my research project. Through his guidance, I improved a great deal in my programing skills, including writing efficient code for data management and analysis of genetic data.

I would like to thank my family and friends for their endless support and encouragement. A special thank you to my grandmother, who I just recently taught how to say good night - she knows less than five English words, but she never fails to encourage me to keep pressing on whenever I call her and tell her about school; and if it happens to be at night, she ends her conversations with "gu-na-i," to mean "good night," always putting a smile on my face.

Last but not least, I would like to express my gratitude to the Vanderbilt Department of Biostatistics for giving me the opportunity to learn in such a friendly and interactive environment and for providing the resources I needed to make the learning experience smoother and rewarding.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER I

Introduction and Background

In practice, researchers collect and analyze data of different types, including continuous and categorical data. To date, various statistical methods have been developed for analysis of ordinal categorical data. Most of the methods are suited for the analysis of ordered categorical outcomes. Examples of models that are commonly used to assess the relationship between an ordinal outcome and continuous or categorical predictors include the cumulative logit model, the continuation-ratio model and the proportional odds model [3]. The treatment of ordinal independent variables is an area that beckons further exploration. Standard regression models treat ordinal predictors as either continuous or as categorical variables. If ordinal predictors are treated as categorical variables, order information is ignored, presumably resulting in loss of power. If treated as continuous variables, a linear relationship between the outcome and predictor is assumed which may not be desirable.

Alternative methods have been proposed to handle ordinal predictors. Walter et al [14] describe a coding scheme that can be used to define contrasts in the dependent variable between successive levels of the predictor, or to identify critical threshold values of the predictors at which significant changes occur in the response. Other methods specific to cases where both the outcome and the predictor are ordinal include the use of splines to impose monotonicity in transformed variables [11], isotonic regression [4], use of latent variables, joint modeling of the ordinal predictor (X) and the outcome (Y) conditional on other covariates (Z) [6, 2, 10], Kendalls partial tau [8], an extension of Kendalls partial tau to multivariable Z [7] and stratifying data according to Z and then computing weighted averages of stratum-specific measures of association between X and Y [13, 5, 1]. As discussed by Li and Shepherd [9], these methods have limitations. For instance, the use of splines requires the specification of number and location of knots, the use of latent variables require the specification of a distribution for the latent variable, and the implementation of some of these methods require grouping of continuous or multivariable Z into strata, which may lead to loss of information due to the arbitrary generation of cut-offs on Z. These limitations hence motivated the development of conditional ordinal by ordinal tests (COBOT) outlined in the subsection that follows.

Conditional Ordinal by Ordinal Test (COBOT)

Li and Shepherd [9] introduced a method for testing the association between two ordinal variables, X and Y, while adjusting for categorical or continuous covariates, **Z**. They developed test statistics that can be used to test for association between X and Y by fitting separate multinomial models of X and Y given

Z. The motivation is that these two conditional distributions will be independent if there is no relationship between X and Y conditional on **Z**. Their first test statistic is based on the comparison of the observed joint distribution between X and Y with their expected distribution under the null of conditional independence, while accounting for order information in Y and X. Similarly, the second test statistic is based on fitting separate models for $P(Y|\mathbf{Z})$ and $P(X|\mathbf{Z})$. Then, the correlation between the residuals from these two models (i.e., cor($Y_{i,res}$, $X_{i,res}$)) is calculated. The third test statistic is a variation of the second approach in which the observed value of each individual (Y_i, X_i) is compared with the distribution of possible values of (Y, X) given covariate $\mathbf{z}_i$. Suppose a random value, (Y_i', X_i'), from subject $i's$ product distribution is drawn and then compared with the observed value; in the absence of ties, the pair of data points (Y_i, X_i) and (Y_i', X_i') is concordant if $X_i > X_i'$ and $Y_i > Y_i'$, or if $X_i < X_i'$ and $Y_i < Y_i'$, and discordant otherwise. The probabilities of concordance (C_i) and discordance (D_i) can be derived under the null of (Y_i, X_i) and (Y_i', X_i') both following the same product distribution, yielding the third test statistic as the average difference of C_i and D_i across all subjects. Further details on the development and performance of the three test statistics can be obtained in their paper.

For this study, we present an extension of COBOT, in which the outcome is continuous and the predictor is ordinal, while adjusting for other covariates. Specifically, we develop an extension of the residual-based test statistic from COBOT to test for the association between a continuous outcome and an ordinal predictor. Henceforth, we refer to this method as the residual-based conditional continuous by ordinal test (CoCoBOT). In the subsequent sections, we introduce the set-up and basic theory of this method and evaluate its performance particularly in a situation where the ordinal predictor is genotype with three categories.

Motivation for Genome Wide Association Studies

Most genome wide association studies (GWAS) of the association between SNPs and a phenotype such as kidney disease or anti-retroviral (ARV) pharmacokinetics involve the specification of additive, dominant or recessive model during analysis. In some cases, the genotype can be treated as a categorical or factor variable. Consider a bi-allelic marker with alleles A and a. The possible genotypes for this marker are A/A, A/a and a/a. Assuming the effect of an outcome associated with a given genotype is represented by β, an additive model with the genotypes coded as $A/A = 0$, $A/a = 1$, and $a/a = 2$ indicates that the risk of the outcome is increased by β (2β) for genotype A/a (a/a). A recessive model with the genotypes coded as $A/A = 0$, $A/a = 0$, and $a/a = 1$ indicates that two copies of the a allele are required for a β increase in risk of the outcome. Finally, a dominant model with the genotypes coded as $A/A = 0$, $A/a = 1$, and $a/a = 1$ indicates that at least one copy of the a allele is required for a β increase in risk of the outcome.

A categorical model on the other hand models the genotypes using indicator variables. Table 1 below shows an illustration of these four model specifications:

Table 1: Model specifications for genetic associations

Analysis Model	**Genotype**		
	A/A	A/a	a/a
Additive	0	β	2β
Recessive	0	0	β
Dominant	0	β	β
Categorical	0	β_1	β_2

The genotypes of a given marker can be characteristically ordered (0, 1, 2 as in the stated example). The additive model assumes linearity, while the categorical model ignores the order information of the genotypes. In light of this, we seek to apply the residual-based CoCoBOT method to analysis of GWAS data, and compare its performance with the additive, dominant, recessive and categorical models using simulations. In addition, we use all five models to detect SNPs associated with two phenotypes (tenofovir clearance and creatinine clearance) using real data and compare results.

In Chapter II, we present a GWAS of tenofovir clearance and creatinine clearance (CrCl) using the additive model. In Chapter III, we introduce the basic theory of CoCoBOT and a simulation study for genotype data and compare the power of the CoCoBOT, additive, dominant, recessive and categorical models to detect genetic associations. In Chapter IV, we present a GWAS case study of tenofovir clearance and CrCl using the dominant, recessive, categorical and CoCoBOT models in addition to the additive model and compare results. Finally in Chapter V, we summarize the results from the simulation study and case study, and the implications of our findings for future research.

CHAPTER II

Genome-wide Association Study of Tenofovir Pharmacokinetics and Creatinine Clearance in AIDS Clinical Trials Group Protocol A5202

In this section, we present a GWAS study conducted concurrently with this work. The study was motivated by findings from previous studies of HIV-infected patients who are randomized to tenofovir, one of the commonly used antiretroviral (ARV) drugs. In this chapter, we present a review of the study, methods, results and conclusions. For our case study (Chapter IV), we perform the same analyses outlined in the study, except we employ additional models (dominant, recessive, categorical) and also use the residual-based CoCoBOT method. We ultimately compare the results we obtain using all five models.

Introduction

Tenofovir disoproxil fumarate (TDF) is included among recommended first-line regimens for human immunodeficiency virus type 1 (HIV-1) infection [15]. It is converted *in vivo* to tenofovir, which undergoes intracellular diphosphorylation to its active moiety, tenofovir diphosphate [16]. Although generally safe, effective and well tolerated [17-27], some HIV-infected patients prescribed TDF experience declines in creatinine clearance (CrCl) [24, 25, 28-30], particularly in regimens that include an HIV-1 protease inhibitor plus low-dose ritonavir [24, 25, 29]. In AIDS Clinical Trials Group (ACTG) protocol A5202, median change in CrCl from baseline to week 96 decreased by 3 mL/min in subjects who received TDF/emtricitabine with atazanavir/ritonavir, but increased by 5 mL/min in subjects who received TDF/emtricitabine with efavirenz [25]. Discontinuation or dose reduction of TDF/emtricitabine for changes in renal function in A5202 was infrequent [25].

Renal elimination of tenofovir involves glomerular filtration and tubular secretion [31]. Tenofovir entry into proximal tubule cells appears to be mediated by two transporters, solute carrier family 22 member 6 (*SLC22A6*, previously called organic anion transporter 1 (*OAT1*)) and *SLC22A7* (previously called *OAT2*) [32]. Renal tubular secretion is mediated by efflux transporters including ATP-binding cassette, sub-family C, member 4 (*ABCC4*, previously called multidrug resistance protein 4 (*MRP4*)) [32], and possibly *ABCC2* (previously called multidrug resistance protein 2 (*MRP2*)), although the importance of *ABCC2* is uncertain [32]. Declines in CrCl associated with concomitant HIV-1 protease inhibitors may reflect *ABCC4* inhibition, with resultant tenofovir accumulation in proximal tubule cells.

Several candidate gene studies involving HIV-positive patients have suggested associations between single nucleotide polymorphisms (SNPs) and adverse renal effects with tenofovir-containing regimens, although none would have been significant at $P < 0.05$ if corrected for multiple comparisons. A study of 30 patients in France (13 cases and 17 controls) suggested increased risk for proximal renal proximal tubulopathy with an *ABCC2* polymorphism (1249G → A, rs2273697, P = 0.02) [33]. A study of 190 Japanese patients (19 cases and 181 controls) suggested increased risk for renal tubular dysfunction with two polymorphisms in *ABCC2* (-24T → C, rs717620, and 1249 G → A, each P = 0.02) [34]. Analyses of a cohort from Spain, 19 kidney tubular dysfunction cases and 96 controls, suggested increased risk with *ABCC2* -24T → C (P = 0.03) [43], and two polymorphisms in *ABCC10* (rs9349256, P = 0.02; and rs2125739, P = 0.05) [35]. In addition, a study of 30 patients suggested an association between higher peripheral blood mononuclear cell tenofovir diphosphate concentrations and an *ABCC4* polymorphism (3463A → G, rs1751034, $P = 0.04$) [22]. Kidney tubular dysfunction included serum creatinine and/or creatinine clearance differences in some previous reports [33][34] but not in others [29][21]. Previous reports showed limited replication other than perhaps *ABCC2* polymorphisms [33][34].

Here we present two genome-wide association studies (GWAS) based on a cohort of HIV-infected subjects who participated in a prospective randomized clinical trial. The first GWAS considers plasma tenofovir clearance, and the second considers change in estimated CrCl. Within each GWAS we performed preplanned analyses of candidate genes and SNPs. To our knowledge, this is the first GWAS of tenofovir pharmacokinetics, and the first GWAS of tenofovir-related change in renal function. We identified SNPs of potential interest, although none were genome-wide significant after Bonferroni correction. Importantly, polymorphisms previously implicated in tenofovir-associated renal toxicity did not replicate in the present study.

Methods

Study Subjects

AIDS Clinical Trials Group (ACTG) Protocol A5202 (ClinTrials.gov NCT00118898) was a phase IIIb equivalence study of four once-daily regimens for initial treatment of HIV-1 infection. Primary results of A5202 have been previously reported [22, 25]. Briefly, 1,858 HIV-infected subjects were randomized to receive either TDF/emtricitabine (300 mg/200 mg) or abacavir/lamivudine (600 mg/300 mg), with either open-label atazanavir (300 mg) plus ritonavir (100 mg), or efavirenz (600 mg). Protocol-defined evaluations of serum creatinine determinations were performed before entry, at entry, at weeks 4, 8, 16 and 24, and every 12 weeks thereafter until week 96 after the last subject enrolled. Creatinine clearance was calculated using

the Cockcroft-Gault formula based on ideal body weight [37].

Tenofovir assays and plasma sampling

Pharmacokinetic samples for this analysis include those collected between weeks 4 and 24 of A5202. A sparse sampling strategy was designed to collect and measure antiretroviral concentrations in 3 plasma samples per subject. Plasma collection times included a 24-hour post-dose sample followed by an observed-dose sample 3-4 hours post-dose, and another sample 5-15 hours post-dose. The TDF dose of 300 mg is equivalent to 245 mg of tenofovir disoproxil and to 136 mg of tenofovir. Steady-state plasma concentrations of tenofovir were measured using tandem mass spectrometry detection at the ACTG Pharmacology Laboratory at the University of Alabama Birmingham [51].

Pharmacokinetic model development

A total of 2,172 plasma tenofovir determinations from 818 participants were analyzed using a non-linear mixed-effects modeling approach (NONMEM version VII; ICON, Ellicott City, MD). One- and two-compartment disposition models with first-order absorption were tested to determine the pharmacokinetic structural model. The first-order conditional estimation method with interaction (FOCE-I) was used throughout. The final model selected was a two-compartment model that estimated the apparent oral and intercompartmental clearances, and volumes of distribution of the central and peripheral compartments. Because concentration data were lacking in the absorption phase, the absorption rate constant was fixed to 1 h-1 based on previous data [61]. Exponential errors with log-normal distribution were used for intersubject variability of pharmacokinetic parameters. A proportional error model was assigned to the residual variability. The final structural pharmacokinetic model was assessed by successful convergence and goodness-of-fit plots. Individual Bayesian estimates of oral clearance values of tenofovir were estimated from the final structural model.

Identifying genetic polymorphisms

Consent for DNA testing was obtained under ACTG protocol A5128 [38]. Of the 1,858 subjects with clinical data, 1,356 consented to A5128, including 677 randomized to tenofovir-containing regimens (350 with efavirenz, 327 with atazanavir/ritonavir) and 679 randomized to abacavir-containing regimens (349 with efavirenz, 330 with atazanavir/ritonavir). Genome-wide genotype data on 1,221 subjects from the Illumina Human-1M-Duo platform were available from a separate immunogenomics project [39]. The Vanderbilt Institutional Review Boards and the ACTG approved this use of genotype data. Genetic data management and association analyses were performed with PLINK version 1.07 [40].

For pharmacokinetic association analyses, we considered a subset of candidate SNPs of potential relevance to tenofovir [33-35, 41-48]. Initially, 25 candidate SNPs in 11 genes suggested to affect tenofovir renal elimination were identified using the pharmGKB database (*ABCB1, ABCC10, ABCC2, ABCC4, AK2, AK3, NME1, SLC22A6, SLC22A8* and *SLC22A11*) [55]. Of the 25 SNPs, 15 were in our genotype data. Proxies were identified for 3 of the other 10 (a total of 15 proxies) through SNP annotation and proxy (SNAP) search [54], based on a pairwise r^2 threshold of 0.8, providing a total of 30 SNPs (representing 18 candidate SNPs) for this pharmacokinetic analysis. For separate analyses we also considered all available SNP data between the transcription start and end positions of these 11 genes, based on the human reference genome hg18/NCBI36 [52]. We identified 594 such SNPs: 110 in *ABCB1*, 12 in *ABCC10*, 38 in *ABCC2*, 378 in *ABCC4*, 8 in *AK2*, 18 in *AK3*, 6 in *NME1*, 6 in *SLC22A6*, 14 in *SLC22A8*, 4 *SLC22A11*, and 0 in *NME1*.

For CrCl association analyses, we identified 91 SNPs previously associated with any renal phenotype at $P < 1.0 \text{ x } 10^{-5}$ in any cohort, as posted to the NHGRI GWAS Catalog [49]. Trait search terms used to survey the GWAS Catalog were "chronic kidney disease", "chronic kidney disease and serum creatinine levels", "creatinine levels", "end-stage renal disease", "glomerulosclerosis", "IgA nephropathy", "nephropathy", "nephropathy (idiopathic membranous)", "nephrotic syndrome (acquired)", "renal function and chronic kidney disease", "renal function-related traits (BUN)", "renal function-related traits (eGRF creatinine)", "renal function-related traits (sCR)" and "renal function-related traits (urea)". Of the 91 SNPs, 43 were available in our genotype data. Proxies were identified for 34 of the other 48 SNPs (a total of 169 proxies) as described above, providing a total of 212 SNPs (representing 77 candidate SNPs) for CrCl association analysis.

Quality control of genetic data

Quality control (QC) of genetic data was done using PLINK version 1.07 [40]. Before QC, 546 subjects met inclusion criteria for the pharmacokinetic (1,212 for CrCl) association analyses as described in the statistical analyses section, and with 1.2 million SNPs. We excluded subjects with greater than 2% missing genotypes, extreme heterozygosity ($|F| > 0.1$), for duplicates or relatedness ($\hat{\pi} > 0.125$), or sex mismatch. We excluded SNPs that deviated from Hardy-Weinberg equilibrium within principal component (PC)-derived race/ethnicity groups ($P < 10^{-6}$), with less than 98% genotyping efficiency, and minor allele frequency (MAF) $< 5\%$. Quality control was performed for the combined group (African Americans, European Americans and Hispanic Americans, hereafter called White, Black, and Hispanic, respectively) and separately for each race/ethnicity group.

We generated PCs to infer and adjust for genetic ancestry. Before generating PCs we removed non-autosomal SNPs, and SNPs from two regions of high linkage disequilibrium (LD), 25.5 MB to 33.5 MB spanning the MHC region on chromosome 6, and from 8.0 MB to 12.0 MB on chromosome 8, because these

regions are not removed by PLINKs indep-pairwise pruning method. Subsets of SNPs in low pairwise LD ($r^2 < 0.2$) were used to generate PCs using smartpca in EIGENSTRAT [59]. Race/ethnicity was derived by analyzing these samples in concert with HapMap 3 samples from 11 populations [39].

Pharmacokinetic association analyses

Pharmacokinetic association analyses included only subjects that had been randomized to tenofovir-containing arms and had available clinical, pharmacokinetic, and genotype data. Multivariable linear regression models were fit. Analyses were performed on all subjects as a combined group, and separately in each group (White, Black, and Hispanic) based on PCs. Meta-analysis of the three stratified models was also performed.

Tenofovir clearance, estimated from pharmacokinetic models as described above, was regressed on genotype, adjusting for sex, age, body mass index (BMI), concomitant efavirenz versus atazanavir/ritonavir, and baseline CrCl. In the combined analysis, we also adjusted for the first two PCs and self-reported race coded as White, Black, Hispanic, and other. For QC we tested for the known genome-wide association between UGT1A1 SNPs and baseline plasma bilirubin concentration [60].

Creatinine clearance association analyses

Creatinine clearance association analyses included subjects randomized to TDF or abacavir arms, and with available clinical and genotype data. All determinations within 200 days after randomization were included in analyses; 200 days was chosen as 6 months of follow-up with a grace period after looking at the timing of creatinine measurements in the database. At least two CrCl determinations after baseline were available from 91% of subjects.

The TDF and abacavir treatments are expected to result in different progression of creatinine clearance over time; we were interested in identifying genes that can influence such a difference. In other words, there is an expected interaction between treatment and time on CrCl, and we wanted to test if the interaction differs between individuals who carry different genotypes at a genetic marker. We addressed this using a one-degree-of-freedom test for the significance of a three-way interaction among time, treatment, and genotype. We performed analysis using a generalized least squares regression model with compound symmetric correlation structure, with creatinine clearance (hereafter called the *time-dependent CrCl change*) as the outcome and including all main, two-way and three-way interactions of time, treatment arm (TDF or abacavir), and genotype. In the model we also adjusted for sex, age, BMI, self-reported race, concomitant antiretroviral (efavirenz or atazanavir/ritonavir), baseline CrCl, and the first two PCs. Baseline CrCl was the value at randomization (i.e., day 0); for subjects without data on day 0, we used the first available value before

randomization (preferred) or after randomization. Three additional models, stratified by PC-inferred race, were fit in the same manner as above, but without adjusting for self-reported race and PCs. Meta-analysis of regression output from the stratified models was also done. For CrCl we evaluated the 212 candidate GWAS Catalog SNPs.

We repeated the analyses as described above, but using as a phenotype change in CrCl from baseline to 6 months (defined as the value closest to day 183 $\pm$ 30 days). We hereafter call this the 6-month CrCl change. This cut-off is based on reported time to change in creatinine, much of which is apparent within the first 6 month of initiating TDF-containing regimens. Time was excluded in this model, and the p-value for two-way interaction between genotype and treatment arm (TDF or abacavir) was calculated.

The finding from [25] suggested an interaction between TDF and atazanavir/ritonavir (ATZ/r) or efavirenz (EFV). Therefore, we repeated the analyses for 6-month CrCl change as described above, with the inclusion of the interaction between treatment arm and concomitant ARV. In addition, we repeated the TDF clearance analyses as described above, but evaluated the combined effect of genotype and genotype-EFV or ATZ/r interaction on TDF clearance using a likelihood ratio test. All analyses were repeated using baseline plasma bilirubin concentration as a positive control phenotype.

Statistical analyses were done using PLINK version 1.07 [40] and R version 3.0.1. Meta-analyses were performed in PLINK, and the random effects p-values reported. Analysis scripts are available upon request. Except where indicated otherwise, we used Bonferroni correction to determine significance thresholds, with $P < 5.0 \times 10^{-8}$ for genome-wide analyses, and 0.05 divided by number of SNPs evaluated in each candidate gene or SNP analysis.

Results

Study subjects and genetic data

Characteristics of subject in the tenofovir and abacavir arms are shown in Table 2. Randomization provided similar distributions of sex, self-reported race/ethnicity, concomitant antiretrovirals, age, BMI, and baseline creatinine clearance between arms. Subjects were predominantly male, and approximately 50% were Black or Hispanic. Figure 1 describes data management and QC steps in the combined group analyses. After QC, for pharmacokinetic analyses there were 501 subjects and approximately 890,000 SNPs. For creatinine clearance analyses there were 1096 subjects (548 randomized to TDF-containing regimens) and approximately 840,000 SNPs.

Table 2 Baseline characteristics of participants

Variable	TDF/FTC (n = 501)	ABC/3TC (n = 548)
Male; No. (%)	434 (87)	472 (86)
Self-reported race/ethnicity; n (%)		
White	243 (49)	243 (44)
Black	149 (30)	189 (36)
Hispanic	100 (20)	102 (18)
Other	9 (2)	14 (2)
Concomitant antiretroviral; n (%)		
Atazanavir	246 (49)	269 (49)
Efavirenz	255 (51)	279 (51)
Age in years; median (IQR)	39.0 (31.0, 45.0)	38.0 (31.0, 45.0)
BMI in kg/m^2; median (IQR)	24.8 (22.3, 27.9)	24.8 (22.3, 27.8)
Baseline CrCl in mL/min; median (IQR)	116.0 (99.8, 135.5)	116.6 (99.2, 138.0)

TDF/FTC = tenofovir disoproxyl fumarate with emtricitabine; ABC/3TC = abacavir with lamivudine; n = number; kg/m^2 = kilogram per square meter; mL/min = milliliter per minute; BMI = body mass index; CrCl = creatinine clearance; IQR = interquartile range.

Pharmacokinetic association analyses

For pharmacokinetic association analyses, no SNP was significantly associated with tenofovir clearance after Bonferroni correction. Table 3 shows the 20 SNPs with the smallest p-values in meta-analyses based on GWAS, candidate SNPs, and candidate genes. Results of similar analyses, but for all subjects adjusting for PC-derived ancestry, and separately within each race/ethnicity group, are provided in Appendix A.

Of 30 candidate SNPs evaluated, 15 (50%) were in *ABCC4*. A LocusZoom plot of this SNP position ± 500 KB was used to investigate the surrounding region. As shown in Figure 2, the SNP with the lowest p-value in this region was in *CLDN10* (rs12866697), not in *ABCC4*. For QC, log-transformed baseline bilirubin concentration was analyzed as a positive control phenotype with a known genotype association [60]. Multiple UGT1A1 SNPs were associated with baseline bilirubin (P = 2.2 x 10-11 for UGT1A1 rs887829), confirming the ability of our analyses to detect true associations.

The analyses in which the effect of genotype and genotype-EFV or ATZ/r interaction on TDF clearance were evaluated yielded similar results to what was initially obtained. No SNPs were genome-wide significant, and the SNPs with the smallest p-values were the same ones seen in the initial analyses (results not shown). Use of baseline plasma bilirubin concentration as a positive control in these analyses, again confirmed our ability to detect true associations.

Table 3 Meta-analysis Results of Pharmacokinetic Associations (top 20 SNPs)

GWAS			**Candidate SNPs (30 SNPs tested)**				**Candidate Genes (594 SNPs tested)**			
SNP	CHR	P	SNP	CHR	gene	P	SNP	CHR	gene	P
GA032783	17	1.4e-07	rs3818486	13	ABCC4	0.03586	rs3847258	9	AK3	0.02925
rs359770	18	4.0e-06	rs2185631	6	ABCC10	0.04814	rs16921966	9	AK3	0.03278
rs359769	18	4.1e-06	rs6421690	11	SLC22A11	0.07005	rs3818486	13	ABCC4	0.03586
rs438697	1	8.2e-06	rs4148486	13	ABCC4	0.08971	rs10798924	1	AK2	0.03672
rs1833170	16	8.4e-06	rs2389204	13	ABCC4	0.107	rs12429339	13	ABCC4	0.03899
rs17199679	9	9.3e-06	rs7924450	11	SLC22A11	0.121	rs11591185	1	AK2	0.04639
rs6429839	1	1.2e-05	rs11231803	11	SLC22A11	0.1235	rs2185631	6	ABCC10	0.04814
rs4662167	1	1.2e-05	rs3818493	13	ABCC4	0.1287	rs2268691	1	AK2	0.05291
rs7829911	8	1.5e-05	rs9349256	6	ABCC10	0.1426	rs1611822	13	ABCC4	0.05928
rs1165176	6	2.0e-05	rs2274405	13	ABCC4	0.171	rs17268129	13	ABCC4	0.05975
rs1420040	16	2.0e-05	rs7331142	13	ABCC4	0.1771	rs1751033	13	ABCC4	0.07184
rs765285	6	2.1e-05	rs9524827	13	ABCC4	0.1816	rs9561765	13	ABCC4	0.07938
rs1165177	6	2.1e-05	rs9394952	6	ABCC10	0.2	rs1564351	13	ABCC4	0.08005
rs1185569	6	2.1e-05	rs1045642	7	ABCB1	0.2381	rs7330330	13	ABCC4	0.08221
rs4319926	2	2.2e-05	rs2273697	10	ABCC2	0.2803	rs4148487	13	ABCC4	0.08363
rs4442993	2	2.2e-05	rs717620	10	ABCC2	0.2872	rs4148442	13	ABCC4	0.08676
rs563189	1	2.5e-05	rs2125739	6	ABCC10	0.3188	rs4148451	13	ABCC4	0.08676
rs10503961	8	2.5e-05	rs4148477	13	ABCC4	0.3348	rs1729745	13	ABCC4	0.08888
rs7144413	14	2.7e-05	rs4148478	13		0.3348	rs4148486	13	ABCC4	0.08971
rs11844480	14	2.8e-05	rs17222723	10		0.4265	rs9524858	13	ABCC4	0.09165

CHR = Chromosome; SNP = SNP identifier; P = P-value. Significance levels were approximately 5×10^{-8} for the genome-wide analyses, 0.002 for the subset of 30 SNPs and 8.4×10^{-5} for the subset of 594 SNPs.

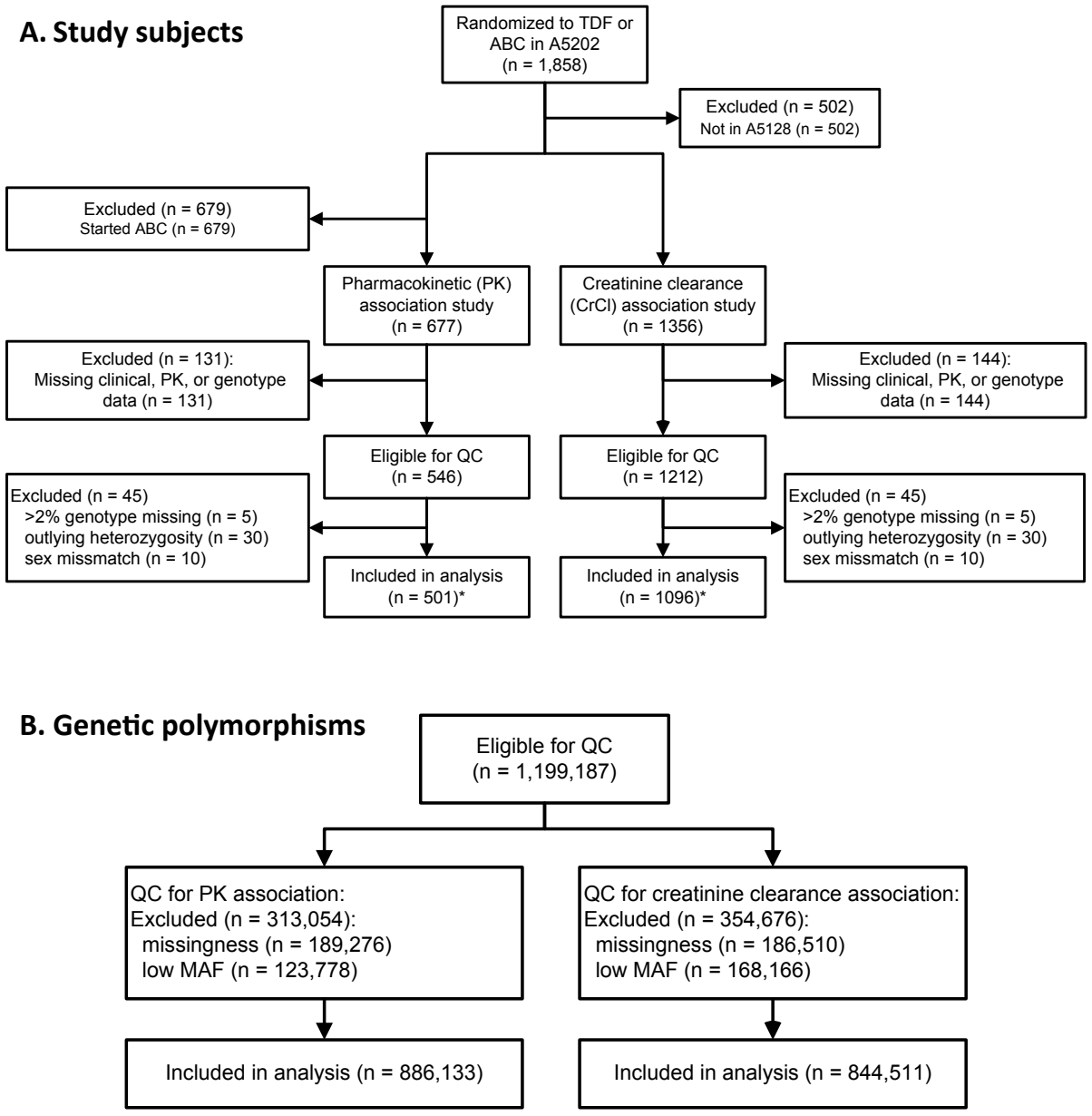

Figure 1 Disposition of study subjects and SNPs through the data management and QC process. Top panel is the disposition of study subjects. The number of subjects included in PK and CrCl association analyses varied depending on the SNPs included in the analysis, with a median (IQR) of 501 (500 to 501) PK analysis subjects, and 1039 (1038 to 1040) CrCl analysis subjects. Bottom panel is the disposition of genetic polymorphisms.

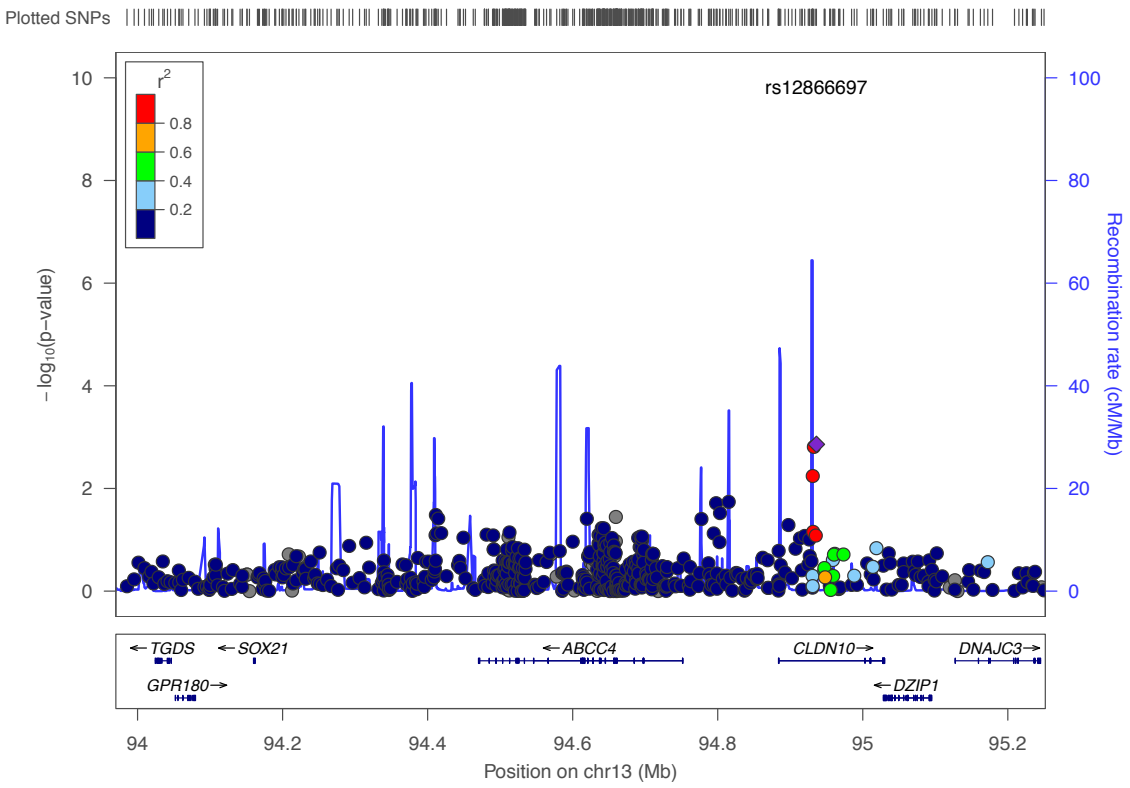

Figure 2 LocusZoom plot of ABCC4 gene region for association with tenofovir pharmacokinetics by meta- analysis. The region of ABCC4 (±500 KB) is shown. Genes in the region are shown at the bottom. Filled circles represent p-values for SNPs in our data. The lowest p-value SNP in this region, rs12866697, is represented by the purple diamond. Markers are color coded to represent their degree of correlation (r^2) with rs12866697 as estimated internally by LocusZoom using the hg18/HapMap Phase II CEU genome build. The blue lines correspond to the recombination rate [56].

Creatinine clearance association analyses

In the time-dependent CrCl change analyses, no SNP was significantly associated after Bonferroni correction. The 20 SNPs with the lowest p-values in GWAS and candidate SNP meta-analyses are shown in Table 4. Results of similar analyses, but for all subjects adjusting for PC-derived ancestry, and separately within each race/ethnicity group, are provided in Appendix A.

In the analysis involving the entire population, rs1751036 in *ABCC4* was among the top 20 SNPs for time-dependent CrCl change ($P = 2.4 \times 10^{-5}$, Table 13). A LocusZoom plot of the *ABCC4* region (± 500 KB) is shown in Figure 3. Twelve *ABCC4* SNPs in LD with rs1751036 at $r^2 > 0.8$ were associated with time-dependent change in creatinine clearance at unadjusted $P < 0.01$, and are listed in the figure legend.

Table 4 Meta-analysis Results of Creatinine Clearance Associations (top 20 SNPs)

GWAS			Candidate SNPs (from GWAS catalogue)			
SNP	CHR	P	SNP	CHR	Gene	P
rs121882	5	1.24e-05	rs10941692	5	Intergenic	0.02643
rs4243086	15	1.45e-05	rs7805747	7	PRKAG2	0.02802
rs9602954	13	1.51e-05	rs12520150	5	Intergenic	0.04025
rs3914576	15	1.96e-05	rs9473932	6	Intergenic	0.05142
rs419129	6	2.54e-05	rs2082424	19	CEP89	0.05577
rs17039196	4	2.91e-05	rs7246178	19	CEP89	0.05923
rs17622266	7	3.03e-05	rs660895	6	Intergenic	0.07165
rs6963950	7	3.04e-05	rs17272197	19	SLC7A9	0.147
rs7002294	8	3.97e-05	rs1133029	20	Intergenic	0.1493
rs7034027	9	4.21e-05	rs653178	12	ATXN2	0.1501
rs2273289	1	4.32e-05	rs2057291	20	GNAS	0.1516
rs10758871	9	4.45e-05	rs6677604	1	CFH	0.1602
rs1976756	15	4.54e-05	rs4664308	2	PLA2R1	0.2388
rs7033976	9	4.57e-05	rs16902083	5	HCN1	0.2478
rs3750494	9	4.69e-05	rs11705804	3	Intergenic	0.2497
rs749074	14	5.54e-05	rs1556751	9	PIP5K1B	0.2523
rs4524177	20	5.61e-05	rs11864909	16	PDILT	0.2545
rs2682621	12	5.92e-05	rs3925075	16	Intergenic	0.2577
rs11067378	12	6.43e-05	rs16853741	3	MECOM	0.2661
rs12701851	7	6.65e-05	rs2151421	9	PIP5K1B	0.2759

Significance levels: approximately 5×10^{-8} for GWAS; 0.0002 for the subset of 212 SNPs.

In analyses by population, among African Americans, considering 6-month CrCl change, rs3127573 in *SLC22A2* was associated in the candidate SNPs analysis (Table 14, $P = 3.3 \times 10^{-5}$), and was also among the top 20 SNPs in the genome-wide analysis (Table 14). This SNP had the second lowest P-value ($P = 0.0018$) among African Americans in the analysis of the 212 candidate SNPs with time-dependent CrCl change (Table 13), and was among the top 20 SNPs in the 212 candidate SNPs analysis of the combined group with time-dependent CrCl change (Table 13, P = 0.09), and in the genome-wide analysis of combined group with 6-month CrCl change (Table 14, $P = 0.036$).

The analyses of 6-month CrCl change in which the interaction between treatment arm and EFV or ATZ/r was included yielded fairly similar results to those in which this interaction was not included. Except for the same SNP (rs3127573, $P = 4.1 \times 10^{-5}$) in the 212 candidate SNPs analysis of the African American group, no polymorphisms were significantly associated with the outcome.

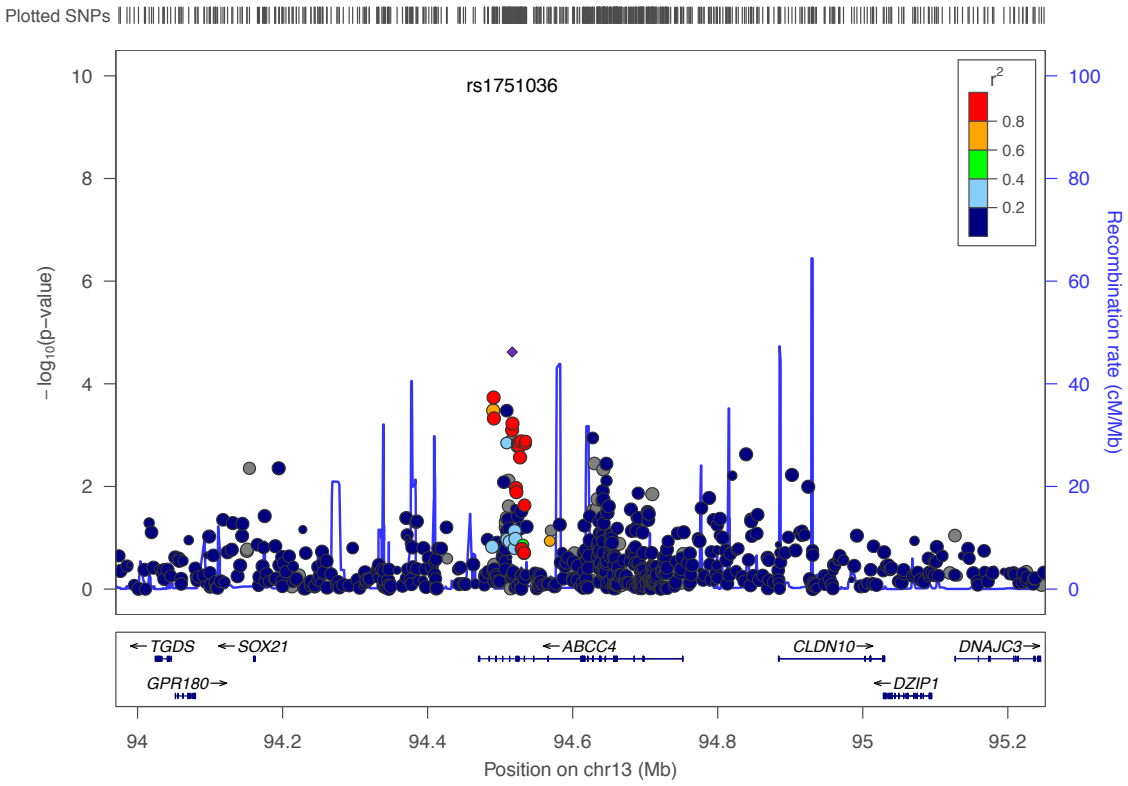

Figure 3 LocusZoom plot of ABCC4 gene region for association with change in creatinine clearance in the entire population. The region of ABCC4 (±500 KB) is shown. Genes in the region are shown at the bottom. Filled circles represent p-values for SNPs in our data. The lowest p-value SNP in this region, rs1751036, is represented by the purple diamond. Markers are color coded to represent their degree of correlation (r^2) with rs1751036 as estimated internally by LocusZoom using the hg18/HapMap Phase II CEU genome build [56]. The 12 SNPs with the lowest p-values are rs7330330, rs7331488, rs4148540, rs2766475, rs1678387, rs1678409, rs1678365, rs1189466, rs1751043, rs943289, rs1189435, and rs1189434.

Discussion

Tenofovir disoproxil fumarate is one of the most extensively prescribed antiretroviral drugs worldwide. The present report describes the first GWAS to investigate associations with tenofovir pharmacokinetics, and the first GWAS to investigate change in CrCl with tenofovir-containing regimens. No polymorphism achieved genome-wide significance ($P < 5.0 \times 10^{-8}$) for association with either tenofovir clearance or change in CrCl. The tenofovir clearance GWAS was complemented by targeted analyses involving 594 SNPs in genes suggested to affect tenofovir disposition, and an even more focused analysis involving 30 candidate SNPs suggested to affect tenofovir disposition. No polymorphism was significant in either of these analyses. Our CrCl GWAS was complemented by a more targeted analyses of 212 SNPs associated with any renal

trait in prior GWAS. Again, no polymorphism was significant in these targeted analyses.

Several aspects of the present study should have favored our likelihood of identifying true-genotype-phenotype associations if present. Both CrCl and tenofovir clearance analysis involved over 500 subjects, far more than were studied in previous genetic association studies of tenofovir renal toxicity [33-35, 43]. The extent of genotype data analyzed far exceeded previous candidate gene analyses [33-35, 43]. Clinical data were from a prospective, randomized clinical trial, which included rigorous quantification of change in creatinine clearance over time, and which showed TDF/emtricitabine with atazanavir/ritonavir to be less favorable in this regard [25, 60, 62]. For CrCl analyses, availability of well matched randomized arms that included abacavir (which is not nephrotoxic) rather than TDF provided leverage to identify genetic associations specific to tenofovir. Substantial numbers of White, Black and Hispanic subjects afforded the opportunity to examine associations both in the combined population, and in each population separately, an approach that has proven valuable in pharmacogenomic analyses of other antiretroviral drugs [60, 62]. Longitudinal models used in our analyses allowed us to capture associations between genotype and change in CrCl over time.

There are several possible reasons for the lack of significant association in the present analyses. With GWAS the threshold for significance after correcting for multiple comparisons is very stringent. However, functional polymorphisms that affect drug disposition and/or pharmacodynamics may be genome-wide significant with modest sample sizes. For example, genetic prediction of abacavir hypersensitivity is genome-wide significant ($P < 5.0 \times 10^{-8}$) with 15 cases and 200 controls [63] and statin response with 85 cases and 90 controls [64]. In addition, we complemented our GWAS with more focused candidate gene/SNPs analyses with less stringent P-value thresholds, which still did not identify significant associations. It is possible that effects of genetic polymorphisms are context dependent, such that they may not have been detected with concomitant efavirenz or atazanavir/ritonavir (the analyses with the EFV or ATZ/r interactions also showed insufficient evidence that the effect of genotype on tenofovir clearance or 6-month CrCl change differed by the EFV or ATZ/r), but could have been apparent with other concomitant antiretrovirals. We cannot exclude the possibility that previously reported associations were spurious, as reported P-values were marginally significant and would not have withstood correction for multiple comparisons even for the few SNPs genotyped.

We considered plasma tenofovir clearance as a phenotype in the present analyses, despite intracellular tenofovir diphosphate being the presumed toxic moiety. Our primary rationale for studying plasma tenofovir clearance was the hypothesis that functional drug transporter gene polymorphisms that affect drug disposition across cell membranes would likely also affect plasma drug concentrations. The lack of significant genetic associations with plasma tenofovir clearance thus reinforces the lack of associations with CrCl.

Within each analysis, there was some overlap for a few SNPs, especially in the combined and meta-

analyses. Although not significant, most of the top SNPs in pharmacokinetic candidate SNP analyses were in *ABCC* genes , which have been shown to play a role in mechanisms of tenofovir clearance and creatinine clearance among patients receiving tenofovir [31-35]. A previous study of the association of *ABCC10* polymorphisms with kidney tubular dysfunction (KTD) identified an association between rs9349256 (odds ratio = 2.3, P = 0.02) and rs2125739 (OR = 2.0, $P = 0.05$) and tubular dysfunction [35]. In our study, rs9349256 was among the top 20 (of 30) candidate SNPs evaluated in the combined, African American and European group analyses, and was among the top 20 (of 594) candidate SNPs evaluated in the European population ($P = 0.10$), which also suggests the potential role of this SNP in tenofovir clearance.

Although none of the SNPs in the *ABCC* family were significant at the genome-wide or candidate SNP level, a SNP near *ABCC4*, rs12866697 located in *CLDN10*, was the top SNP in the *ABCC4* region (± 500 KB). A biological study in mice linked loss of *CLDN10* to hypermagnesemia and nephrocalcinosis because this gene is involved in paracellular sodium permeability [57], but the relevance of this to tenofovir is not apparent.

Only the sensitivity analysis of the 212 candidate SNPs identified SNP rs3127573 as being significant in the CrCl associations. However, this SNP appeared among the top 20 SNPs in some of our genome-wide and candidate SNPs analyses as outlined in the results section. Previous studies have shown the potential role played by *SLC22A6* and *SLC22A7* in the renal proximal tubules [32, 46]. Furthermore, a study of the association of *SLC22A2* polymorphisms with phenotypes of net tubular creatinine secretion in which SNP rs3127573 was one of the two SNPs genotyped in patients with end-stage renal disease found a positive association between end-stage renal disease and SNP rs3127573: odds ratios [95% CI] 1.39 [1.16-1.67] [58]. Our results are in the same direction as this finding and hence affirm the association of some SNPs in *SLC22A2* with renal phenotypes.

There were limitations to the present study. Renal toxicity associated with tenofovir in A5202 was modest, so did not include subjects representing extreme phenotypes. The present analysis focused on change in CrCl as the primary phenotype, but it is possible that other markers of renal tubular function are more affected by genotype. The sample size within each PC-derived race/ethnicity was small, although studies have reported significance with even fewer individuals [33-35].

In summary, we did not identify significant genetic associations with plasma tenofovir clearance of change in CrCl among patient randomized to TDF-containing regimens in A5202. Further research is warranted to determine whether previously suggested genetic associations with tenofovir-associated renal tubular injury depend on context, such as specific concomitant medication.

CHAPTER III

New Method: Residual-based Conditional Continuous by Ordinal Test (CoCoBOT)

Motivation

Some of the previously discussed methods for analysis of ordinal data either assume linearity, or ignore order information of the categorical variable. On the other hand, the residual-based CoCoBOT method allows a monotone relationship between a continuous outcome (Y) and an ordered categorical predictor (X), after adjusting for other covariates ($\mathbf{Z}$) that can be continuous or categorical. This is accomplished by fitting separate regression models of Y on $\mathbf{Z}$ (e.g. linear regression), and X on $\mathbf{Z}$ (e.g. proportional odds) and then a residual-based test statistic is constructed. Specifically, the correlation between probability-scale residuals from the two models is assessed. In the subsections that follow, we develop the theory of residual-based CoCoBOT by adapting the methods for the second test statistic described in the COBOT paper by Li and Shepherd [9]. From COBOT, given an ordinal outcome, Y, an ordinal predictor, X and covariates, $\mathbf{Z}$, the residual-based test statistic is constructed by fitting the models for $P(Y|\mathbf{Z})$ and $P(X|\mathbf{Z})$, obtaining the residuals from each, i.e., $Y_{i,res}$ and $X_{i,res}$, respectively, and then testing for the correlation between $Y_{i,res}$ and $X_{i,res}$.

Basic Theory

Definition of Residual For an Ordinal Outcome

Suppose an individual has an observed outcome $Y = y$ and inputs $\mathbf{Z} = \mathbf{z}$. A linear regression of Y on $\mathbf{Z}$ produces the fitted value, $\hat{y} = E(Y|\mathbf{z})$. The observed minus expected residual is defined as the difference between the observed response value, y, and the predicted value, $\hat{y}$, i.e., $(y - \hat{y})$. The development of the residual for an ordered categorical outcome, X, is based on the definition of the residual from linear regression as the expectation of a random variable, $(x - X_{fit})$, where $X_{fit} \sim X|\mathbf{z}$, and $X|\mathbf{z}$ is the distribution of possible outcome values given $\mathbf{z}$. Since it is not possible to calculate $\mathrm{E}(x - X_{fit})$ when X is an ordinal variable (with s categories), for an individual i with outcome $X_i = x_i$, it is plausible to compare x_i and $X_{i,fit}$ with respect to whether x_i is at a lower or higher level than $X_{i,fit}$. The probability for x_i to be higher than $X_{i,fit}$ is $p_{i,high} = P(x_i > X_{i,fit}) = \gamma_i^{x_i - 1}$, where $\gamma_i^j = P(X \leq j|\mathbf{Z}_i = \mathbf{z}_i)$, for $j = 1, \ldots, s$. The probability for x_i to be lower than $X_{i,fit}$ is $p_{i,low} = P(x_i < X_{i,fit}) = 1 - \gamma_i^{x_i}$. The probability for x_i to tie with $X_{i,fit}$ is

$P(x_i = X_{i,fit}) = p_i^{x_i}$. Scores are then assigned to these three events such that 1 = higher, -1 = lower, and 0 = tie. The expected score, $X_{i,res} = p_{i,high} - p_{i,low}$, is a function of data $(X_i, \mathbf{Z}_i)$ and model parameters $\boldsymbol{\theta}^X$.

Definition of Test Statistic

In this setting, the outcome, Y is continuous while the predictor X is ordinal. Therefore, for subject i, the observed minus expected residuals from the model for $P(Y|\mathbf{Z})$ is $R_i^y = E(y_i - Y_{i,fit})$. The probability-scale residuals can be determined by either assuming a normal distribution or empirically; we show how the residuals can be estimated empirically. Define $F(R_i^y) = \frac{1}{n}\sum_{j=1}^{n} I(R_j^y \leq R_i^y)$ as the empirical cumulative distribution function (cdf) of the R_j^y's. Assuming $A \sim F$, the residual for individual i, of a similar form to that described above for ordinal outcomes, is thus defined as:

$$\begin{aligned} Y_{i,res} &= P_F(A < R_i^y) - P_F(A > R_i^y) \\ &= P_F(A < R_i^y) - (1 - P_F(A \leq R_i^y)) \\ &= 2\frac{1}{n}\sum_{j=1}^{n} I(A \leq R_i^y) - \frac{1}{n}\sum_{j=1}^{n} I(A = R_i^y) - 1 \\ &= 2F(R_i^y) - \frac{1}{n}\sum_{j=1}^{n} I(A = R_i^y) - 1. \end{aligned}$$

On the other hand the residual from the model for $P(X|\mathbf{Z})$ is defined as described above for an ordinal outcome, i.e., for a fitted model with parameter estimates $\hat{\boldsymbol{\theta}}^X$, the residual for subject i is defined as $x_{i,res} = X_{i,res|\hat{\boldsymbol{\theta}}^X}$. The residual-based test statistic is defined as:

$$\begin{aligned} T &= cor(Y_{res}, X_{res}) \\ &= \frac{cov(Y_{res}, X_{res})}{\sqrt{var(Y_{res})var(X_{res})}} \end{aligned}$$

Under the null, the conditional distributions of Y and X given $\mathbf{Z}$ are independent, in which case $cov(Y_{res}, X_{res}) = 0$ and subsequently, $T = 0$. Under the null, $\hat{T}$ converges to zero as the sample size goes to infinity.

Determination of P-value For Test Statistic

Two approaches for obtaining the distribution of the test statistic, T, defined above are described in detail in Li and Shepherd [9]. They describe an empirical approach and a large sample approximation for obtaining the distribution of $\hat{T}$, and show via simulation studies that under correctly specified models, high power is achieved regardless of which approach is used to compute the p-value for the test statistic. For our study,

we will use the large sample approximation approach to determine the p-value for T. Briefly, the asymptotic distribution approach relies on the M-estimation method [12], in which estimates of a vector of p parameters, $\boldsymbol{\theta}$, are obtained by finding the solution to the vector equation $\sum_{i=1}^{n} \boldsymbol{\Psi}_i(\boldsymbol{\theta}) = 0$, where $\boldsymbol{\Psi}_i = \boldsymbol{\Psi}(Y_i, X_i, \boldsymbol{Z_i}; \boldsymbol{\theta})$, is a (p x 1)-function that does not depend on i or n. Provided $\boldsymbol{\Psi}_i(\boldsymbol{\theta})$ is sufficiently smooth, i.e., has continuous derivatives up to some desired order over some domain, and $\boldsymbol{\theta}$ has fixed dimension, as n $\longrightarrow \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, V(\boldsymbol{\theta})),$$

where $V(\boldsymbol{\theta}) = A(\boldsymbol{\theta})^{-1}B(\boldsymbol{\theta})[A(\boldsymbol{\theta})^{-1}]', A(\boldsymbol{\theta}) = E[-\frac{\partial}{\partial\theta}\boldsymbol{\Psi}_i(\boldsymbol{\theta})]$, and $B(\boldsymbol{\theta}) = E[\boldsymbol{\Psi}_i(\boldsymbol{\theta})\boldsymbol{\Psi}_i(\boldsymbol{\theta})']$. If $T = g(\hat{\boldsymbol{\theta}})$ is a smooth function of $\hat{\boldsymbol{\theta}}$, then from the delta method,

$$\sqrt{n}[g(\hat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})] \xrightarrow{d} N(\mathbf{0}, \sigma^2),$$

where $\sigma^2 = [\frac{\partial}{\partial\theta}g(\boldsymbol{\theta})]V(\boldsymbol{\theta})[\frac{\partial}{\partial\theta}g(\boldsymbol{\theta})]'$. The estimators of $A(\boldsymbol{\theta})$, $B(\boldsymbol{\theta})$ and $\frac{\partial}{\partial\theta}g(\boldsymbol{\theta})$ are:

$$\widehat{A(\boldsymbol{\theta})} = \frac{1}{n}\sum_{i=1}^{n}[-\frac{\partial}{\partial\theta}\boldsymbol{\Psi}_i(\hat{\boldsymbol{\theta}})],$$
$$\widehat{B(\boldsymbol{\theta})} = \frac{1}{n}\sum_{i=1}^{n}[\boldsymbol{\Psi}_i(\hat{\boldsymbol{\theta}})\boldsymbol{\Psi}_i(\hat{\boldsymbol{\theta}})'],$$
$$\widehat{\frac{\partial}{\partial\theta}g(\boldsymbol{\theta})} = \frac{\partial}{\partial\theta}g(\hat{\boldsymbol{\theta}}).$$

Under the null, if $g(\boldsymbol{\theta}) = 0$, the p-value can be approximated as $2\boldsymbol{\Phi}\big(\frac{-|T|}{\hat{\sigma}\sqrt{n}}\big)$, where $\boldsymbol{\Phi}$ is the cdf of the standard normal distribution.

Definition of Estimating Function, $\boldsymbol{\Psi}(\boldsymbol{\theta})$

We now define the estimating equations used to calculate the p-value described in the previous subsection. The parameter vector, $\boldsymbol{\theta}$, has the form, $\boldsymbol{\theta} = (\boldsymbol{\theta}^Y, \boldsymbol{\theta}^X, \boldsymbol{\theta}^T)$, where $\boldsymbol{\theta}^T$ is the parameter vector for the residual-based test statistic. The resultant estimating function, $\boldsymbol{\Psi}(Y_i, X_i, \boldsymbol{Z_i}; \boldsymbol{\theta})$ has the form:

$$\boldsymbol{\Psi}_i(\boldsymbol{\theta}) = \begin{cases} Y_i - \boldsymbol{Z_i}\boldsymbol{\theta}^Y \\ (Y_i - \boldsymbol{Z_i}\boldsymbol{\theta}^Y)\boldsymbol{Z_i} \\ \frac{d}{d\boldsymbol{\theta}^X} l_X(X_i, \boldsymbol{Z_i}; \boldsymbol{\theta}^X) \\ \boldsymbol{\psi}(Y_i, X_i, \boldsymbol{Z_i}; \boldsymbol{\theta}^T) \end{cases}$$

where $\boldsymbol{Z_i}\boldsymbol{\theta}^Y = Y_{i,fit}$, the fitted value from a linear model with parameters $\boldsymbol{\theta}^Y$ and l_X is the log-likelihood function of the model for $P(X|\mathbf{Z})$, with parameters $\boldsymbol{\theta}^X$. $E[\frac{d}{d\boldsymbol{\theta}^X} l_X(X_i, \boldsymbol{Z_i}; \boldsymbol{\theta}^X)] = 0$, since it is the expected value of a score function. $\boldsymbol{\theta}^T = (w_1, w_2, w_3, w_4, w_5)$, where $w_1 = E(Y_{i,res})$, $w_2 = E(X_{i,res})$, $w_3 = E(Y_{i,res}X_{i,res})$, $w_4 = E(Y_{i,res}^2)$ and $w_5 = E(X_{i,res}^2)$. The resultant estimating function is:

$$\boldsymbol{\psi}(Y_i, X_i, \boldsymbol{Z_i}; \boldsymbol{\theta}^T) = \begin{cases} Y_{i,res} - w_1 \\ X_{i,res} - w_2 \\ Y_{i,res}X_{i,res} - w_3 \\ Y_{i,res}^2 - w_4 \\ X_{i,res}^2 - w_5 \end{cases}$$

Solving the equation $\sum_{i=1}^{n} \boldsymbol{\psi}(Y_i, X_i, \boldsymbol{Z_i}; \boldsymbol{\theta}^T) = 0$, we have $\hat{w_1} = \frac{1}{n}\sum_{i=1}^{n} y_{i,res}$, $\hat{w_2} = \frac{1}{n}\sum_{i=1}^{n} x_{i,res}$, $\hat{w_3} = \frac{1}{n}\sum_{i=1}^{n} y_{i,res}x_{i,res}$, $\hat{w_4} = \frac{1}{n}\sum_{i=1}^{n} y_{i,res}^2$ and $\hat{w_5} = \frac{1}{n}\sum_{i=1}^{n} x_{i,res}^2$. Let $g(\boldsymbol{\theta}) = cor(Y_{i,res}, X_{i,res}) = \frac{E(Y_{i,res}X_{i,res}) - E(Y_{i,res})E(X_{i,res})}{\sqrt{E[Y_{i,res}^2 - (EY_{i,res})^2]E[X_{i,res}^2 - (EX_{i,res})^2]}} = \frac{w_3 - w_1 w_2}{\sqrt{(w_4 - w_1^2)(w_5 - w_2^2)}}$. Then, under the null, $g(\boldsymbol{\theta}) = 0$ and $T = g(\hat{\boldsymbol{\theta}})$. Given $g(\boldsymbol{\theta}) = (w_3 - w_1 w_2) \times \left[(w_4 - w_1^2)(w_5 - w_2^2)\right]^{-1/2} = num \times den$, the partial derivatives for $g(\boldsymbol{\theta})$ are defined as follows:

$$\begin{aligned} \frac{d}{dw_1} g(\boldsymbol{\theta}) &= -w_2 den - \frac{1}{2} num \times den^3(-2w_1(w_5 - w_2^2)) \\ \frac{d}{dw_2} g(\boldsymbol{\theta}) &= -w_1 den - \frac{1}{2} num \times den^3(-2w_2(w_4 - w_1^2)) \\ \frac{d}{dw_3} g(\boldsymbol{\theta}) &= den \\ \frac{d}{dw_4} g(\boldsymbol{\theta}) &= -\frac{1}{2} num \times den^3(w_5 - w_2^2) \\ \frac{d}{dw_5} g(\boldsymbol{\theta}) &= -\frac{1}{2} num \times den^3(w_4 - w_1^2) \end{aligned}$$

Simulation Study

In this section, we simulate data to evaluate the performance of the residual-based CoCoBOT method, additive, dominant, and recessive models under different data generation specifications, and compare and contrast finding. Two sets of simulations are performed; in the first one, Z is generated conditioned on X and in the second one, X is generated conditioned on Z using the proportinal odds model.

Methods

Data were generated under five gene models including additive, dominant, recessive and two non-linear gene models. The analysis with CoCoBOT involves fitting a proportional odds (PO) model of $X|Z$ in order to calculate the probability-scale residuals. In the first simulation, we generated $Z|X$ and then fit a model of $X|Z$ using PO model. To ensure our Silumation 1 results were not affected by model misspecification, we repeated simulation 2, in which we generated $X|Z$ under the PO model and then fit the a PO model of $X|Z$. Datasets of size N=500 were generated in the following manner:

- Simulation 1:
  - Genotype, X (0, 1, or 2) was generated under Hardy-Weinberg equilibrium (HWE), that is, from a multinomial distribution with expected genotype probabilities, $P(aa) = p^2$, $P(AA) = q^2$ and $P(Aa) = 2pq$, where $q = 1 - p$, and p represents the minor allele frequency (MAF) and was set to 0.5, 0.3, 0.1 or 0.05. The minor allele is denoted with a and the major allele denoted by A.
  - Z was generated from $N(0.5X, 4)$.
  - The phenotype Y was drawn from $N(1 + \eta f(X) + 0.5Z, 1)$. $\eta f(X)$ specifies the relationship between the number of alleles and the phenotype, and was set to be one of the following for $X = 0, 1, 2$:
    - Additive, $f(X) = X$; $\eta = 0.2$
    - Dominant; $f(X) = 0, 1, 1$; $\eta = 0.3$
    - Recessive; $f(X) = 0, 0, 1$; $\eta = 0.3$
    - Non-linear 1; $f(X) = 0, 1, 4$; $\eta = 0.1$
    - Non-linear 2; $f(X) = 0, 3, 4$; $\eta = 0.1$
- Simulation 2:
  - Z was generated from $N(0, 1)$.
  - Genotype, X (0, 1, or 2) was generated from a proportional odds model such that $P(X \leq j|Z) = [1 + \exp(-(\alpha^X + \beta^X Z))]^{-1}$ for $j = 1, \ldots, 3$. For each data generation, the values of α^X and β^X were adjusted accordingly to attain MAFs of 0.5, 0.3, 0.1 and 0.05.
  - The phenotype Y was drawn from $N(1 + \eta f(X) + 0.5Z, 1)$, with $\eta f(X)$ specified as in Simulation 1 above except for the recessive model, η was set to 0.4 instead of 0.3.

To each dataset, we fit a linear regression model including Z and X and assuming the relationship between X and Y conditional on Z was one of the following:

- Additive; linear relationship
- Dominant; X dichotomized with $X = 1, 2$ put in the same group
- Recessive; X dichotomized with $X = 0, 1$ put in the same group
- Categorical; X treated as categorical variable

Finally, analysis using the new approach (residual-based CoCoBOT) was performed, fitting a linear model of Y on Z, a proportional odds model of X on Z, and assessing the correlation between the probability-scale residuals. This was done for each of 1000 simulation replications. Power was compared by computing the proportion of simulation replications that found an association between Y and X conditional on Z with p-value < 0.05.

Results

Simulation 1 results are presented in Table 5. With data generated using an additive model, the residual-based CoCoBOT approach was only slightly less powerful than the properly specified additive model. The CoCoBOT model was slightly more powerful than the categorical and dominant models, and easily outperformed the recessive model at minor allele frequencies < 0.5. On the other hand, with data generated using the dominant model, the CoCoBOT approach was less powerful than the properly specified dominant model, and outperformed the recessive model specification that had the lowest power among all five models. The categorical model performed better than CoCoBOT at MAF $= 0.5$, and was almost as good as the CoCoBOT and additive models at MAF < 0.5. CoCoBOT was slightly more powerful than the additive model but there was no clear winner between the two under this setting.

As expected, the CoCoBOT model was outperformed by the recessive model when data was generated using a recessive model. The categorical model performed better than the CoCoBOT, additive and dominant models under this setting; CoCoBOT outperformed the dominant model and was only slightly less powerful than the additive model. With data generated under the non-linear 1 model, the CoCoBOT model performed better than the dominant and categorical models at MAF $= 0.5$. However, the additive model was the winner at MAF < 0.5. With data generated under the non-linear 2 model, the CoCoBOT model was generally better than the recessive model, fairly similar to the additive model, and slightly worse than the dominant model. Results for data generated under the recessive and non-linear 1 models tended to be similar because the form of the basis function is comparable; (0,0,1) and (0,1,4) respectively. Likewise, the results for data generated under the dominant and non-linear 2 models were fairly similar because the form of the basis function is comparable; (0,1,1) and (0,3,4) respectively.

Table 5 Power of Estimators to Detect Genetic Associations: $Z|X$ generated from multinomial distribution

Data Generation Model	Analysis Model	Minor Allele Frequency(p)			
		0.5	0.3	0.1	0.05
Additive	CoCoBOT	0.848	0.783	0.447	0.286
(0,1,2)	Additive	0.867	0.811	0.478	0.283
	Dominant	0.72	0.717	0.457	0.287
	Recessive	0.723	0.475	0.13	0.074
	Categorical	0.804	0.725	0.378	0.244
Dominant	CoCoBOT	0.638	0.881	0.73	0.484
(0,1,1)	Additive	0.661	0.865	0.722	0.483
	Dominant	0.817	0.917	0.754	0.503
	Recessive	0.142	0.174	0.084	0.074
	Categorical	0.735	0.86	0.652	0.451
Recessive	CoCoBOT	0.616	0.182	0.052	0.047
(0,0,1)	Additive	0.629	0.261	0.06	0.053
	Dominant	0.144	0.096	0.052	0.049
	Recessive	0.816	0.471	0.095	0.073
	Categorical	0.723	0.378	0.067	0.06
Non-linear 1	CoCoBOT	0.86	0.536	0.183	0.128
(0,1,4)	Additive	0.882	0.616	0.22	0.131
	Dominant	0.481	0.379	0.187	0.121
	Recessive	0.903	0.612	0.142	0.079
	Categorical	0.849	0.599	0.189	0.113
Non-linear 2	CoCoBOT	0.879	0.921	0.767	0.528
(0,3,4)	Additive	0.89	0.918	0.765	0.535
	Dominant	0.892	0.933	0.772	0.537
	Recessive	0.474	0.355	0.114	0.072
	Categorical	0.865	0.886	0.686	0.465

Table 6 shows the results from Simulation 2 in which genotype data was generated under the PO model. Except for the difference in the actual values observed, the results are similar to those obtained from Simulation 1; for instance, with data generated under the additive model, the additive model had the highest power; CoCoBOT was slightly less powerful than the additive model, and the recessive model had the lowest power. In addition, under the nonlinear 1 data generation setting, the additive model had the highest power at MAF < 0.5 but was not much better than CoCoBOT which had similar power.

Based on the simulation results, low power is generally observed under low minor allele frequencies (≤ 0.1), in which the probability of the a/a genotype is (≤ 0.01). This is evident especially for the recessive and non-linear 1 data generation models, in which most of the effect of the genotype is due to the "a" allele, and thus power is low for all analysis models.

Table 6 Power of Estimators to Detect Genetic Associations: $X|Z$ generated under a PO model

Data Generation Model	Analysis Model	Minor Allele Frequency(p)			
		0.5	0.3	0.1	0.05
Additive	CoCoBOT	0.637	0.803	0.447	0.349
(0,1,2)	Additive	0.656	0.834	0.478	0.354
	Dominant	0.361	0.744	0.464	0.347
	Recessive	0.437	0.518	0.107	0.11
	Categorical	0.548	0.756	0.392	0.288
Dominant	CoCoBOT	0.373	0.815	0.749	0.45
(0,1,1)	Additive	0.4	0.795	0.769	0.442
	Dominant	0.642	0.874	0.774	0.464
	Recessive	0.077	0.117	0.082	0.073
	Categorical	0.532	0.797	0.691	0.389
Recessive	CoCoBOT	0.177	0.111	0.07	0.058
(0,0,1)	Additive	0.196	0.139	0.093	0.057
	Dominant	0.047	0.06	0.058	0.048
	Recessive	0.316	0.249	0.156	0.084
	Categorical	0.237	0.201	0.124	0.064
Non-linear 1	CoCoBOT	0.579	0.489	0.27	0.187
(0,1,4)	Additive	0.631	0.574	0.343	0.246
	Dominant	0.145	0.335	0.266	0.184
	Recessive	0.676	0.556	0.31	0.207
	Categorical	0.614	0.526	0.314	0.202
Non-linear 2	CoCoBOT	0.672	0.943	0.742	0.49
(0,3,4)	Additive	0.692	0.938	0.746	0.482
	Dominant	0.707	0.944	0.75	0.512
	Recessive	0.171	0.351	0.19	0.202
	Categorical	0.664	0.903	0.654	0.419

Discussion

In conclusion, the CoCoBOT method does not appear to be generally robust for genetic association studies. It seems to struggle particularly under the recessive data generation scenario - the effect here is mainly driven by the a/a genotype, and the probability of observing this genotype is generally low, resulting in low power to detect associations. Except under the truth, the CoCoBOT approach easily outperforms the recessive and dominant models. Thus, *a priori* knowledge should drive the selection of either the dominant or recessive models; otherwise, model misspecifications can lead to significant loss of detection power.

Although CoCoBOT's performance is fairly similar to that of the additive model, the additive model seems to be more robust for genetic data analysis, especially since in real data settings, the data generation model is unknown. The categorical approach generally performs well even in nonlinear scenarios, and in

addition to the additive model, might also be a good choice for detection of SNPs. The low power under low MAF's however, does not seem to be dependent on the analysis model chosen; it is potentially due to the small probability of observing the genotype that is contributing to most of the effect. Finally, the results from Simulation 2, in which genotype was generated under the proportional odds model, suggest that our results from Simulation 1 were not driven by model misspecification since the two simulations led to similar conclusions about the performance of the five methods.

CHAPTER IV

Case Study

In this chapter, we use the ACTG data (introduced and used in Chapter II) to compare the performance of the five analysis models described in Chapter III. We focus on detection of SNPs associated with the two outcomes, tenofovir clearance and 6-month creatinine clearance (CrCl) change, both described in Chapter II. Detailed descriptions of the study subjects, tenofovir assays and plasma sampling procedures, pharmacokinetic model development and quality control of genetic data are given in the Methods section of Chapter II. For all statistical analyses, PLINK version 1.07 and R version 3.0.1 were used. Bonferroni correction was used to determine significance thresholds, with $P < 5.0 \times 10^{-8}$ for genome-wide analyses.

Methods

Tenofovir pharmacokinetic association analyses

Pharmacokinetic analyses of tenofovir clearance included only subjects that had been randomized to tenofovir-containing arms and had available clinical, pharmacokinetic, and genotype data. Analyses were performed on all subjects, and then separately based on PC-derived race (White, Black, and Hispanic).

For each analysis, five multivariable regression models were fit using tenofovir (TDF) clearance as the primary outcome (Y), genotype as the main predictor (X) modeled with additive, dominant, recessive, categorical and CoCoBOT models. All models were adjusted for sex, age, body mass index (BMI), concomitant efavirenz versus atazanavir/ritonavir, and baseline CrCl. In the combined analysis, we also adjusted for the first two PCs and self-reported race/ethnicity coded as White, Black, Hispanic, and other. We also tested for known genome-wide association between *UGT1A1* SNPs and baseline plasma bilirubin concentration.

Creatinine clearance association analyses

The analyses in this subsection included subjects randomized to TDF and abacavir arms, and with available clinical and genotype data. For each analysis, multivariable regression models were fit using the additive, dominant, recessive, categorical and CoCoBOT model specifications. In the analysis of all subjects, 6-month CrCl change, defined in Chapter II, was regressed on genotype and all models were adjusted for sex, age, BMI, self-reported race, concomitant antiretroviral (efavirenz or atazanavir/ritonavir), baseline CrCl, and the first two PCs.

Three additional models stratified by PC-inferred race were separately fit using each of the five model specifications; 6-month CrCl change was regressed on similar variables as in the analysis of all subjects,

with the exclusion of self-reported race and PCs from the models. For quality control, all analyses involving subjects randomized to TDF and abacavir were repeated as described above but using baseline plasma bilirubin as the outcome.

Results

Table 2 shows a detailed summary of study participants' baseline characteristics, and Figure 1 shows a summary of the data management and QC steps employed in the dataset containing all individuals regardless of race/ethnicity. For the TDF pharmacokinetic association analyses, 501 individuals and approximately 890,000 SNPs were included. For creatinine clearance association analyses there were 1039 subjects and approximately 840,000 SNPs.

Tenofovir pharmacokinetic association analyses

Table 7 shows a pairwise correlation matrix of p-values from the combined group analyses under the five different analysis models. There is a strong correlation of p-values between CoCoBOT and the additive model ($\rho = 0.801$); CoCoBOT is also strongly correlated with the dominant model ($\rho = 0.709$) but has low correlation with the recessive model ($\rho = 0.204$).

Table 7 TDF clearance: Correlation matrix of p-values from combined group analysis

	CoCoBOT	**Additive**	**Dominant**	**Recessive**	**Categorical**
CoCoBOT	1				
Additive	0.801	1			
Dominant	0.709	0.733	1		
Recessive	0.204	0.326	0.067	1	
Categorical	0.557	0.648	0.643	0.653	1

The categorical model seems to be fairly correlated with the additive, recessive and dominant models ($\rho \approx 0.650$), but slightly less with CoCoBOT. A very similar pattern of correlation was observed (not shown) in the models stratified by PC-inferred race/ethnicity (African American, Europeans, Hispanics). Quality control analyses with bilirubin as the outcome also yielded results with a similar pattern; Table 16 of Appendix B shows the correlation matrix from the analysis of all subjects.

Table 8 shows the 10 SNPs with the smallest p-values in combined group analysis under each of the five analysis models. The first 4 (3) SNPs in the recessive (categorical) model were genowe-wide significant. Furthermore, three of the SNPs (rs12082252, rs337298, rs16823145), all in chromosome 1, were similar between the two model specifications.

Table 8 TDF clearance: SNPs with the smallest p-values in combined group analysis

CoCoBOT			Additive			Dominant			Recessive			Categorical		
CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P
23	rs12387850	2.80e-07	23	rs12387850	1.26e-06	23	rs12387850	5.63e-07	1	rs12082252	2.28e-10	1	rs337298	7.80e-10
6	rs1141034	2.63e-06	17	GA032783	2.14e-06	19	rs3815748	5.17e-06	1	rs337298	7.58e-10	1	rs12082252	1.69e-09
3	rs4688755	4.28e-06	18	rs4891230	6.46e-06	5	rs6878929	6.96e-06	1	rs16823145	3.01e-08	1	rs16823145	4.87e-08
23	rs6653311	4.88e-06	18	rs9950415	6.76e-06	16	rs12929434	7.09e-06	19	rs263053	4.27e-08	17	GA032783	8.75e-08
3	rs2883057	4.92e-06	9	rs2478851	9.58e-06	4	rs11723812	7.41e-06	3	rs854207	1.51e-07	19	rs263053	3.02e-07
3	rs4688690	4.92e-06	1	rs11165778	9.59e-06	6	rs1141034	8.05e-06	18	rs966634	3.28e-07	9	rs7849259	4.91e-07
3	rs2526751	4.92e-06	13	rs8001616	9.66e-06	15	rs12440239	1.00e-05	1	rs6695258	3.85e-07	1	rs6695258	5.76e-07
3	rs9311446	4.92e-06	4	rs11723812	1.19e-05	14	rs7144413	1.14e-05	21	rs9982266	4.80e-07	2	rs12479145	6.08e-07
11	rs12420080	4.92e-06	2	rs1375178	1.27e-05	4	rs7687008	1.26e-05	1	rs4391664	5.11e-07	1	rs17641977	6.69e-07
3	rs4688758	5.04e-06	17	rs199529	1.41e-05	17	rs1230103	1.65e-05	3	rs1469386	5.14e-07	9	rs11791293	6.81e-07

Table 9 shows the ranks of the unique SNPs from Table 8 under each analysis model, and their respective p-values. SNP rs12387850 was ranked first in the CoCoBOT ($P = 2.80 \times 10^{-7}$), additive ($P = 1.26 \times 10^{-6}$) and dominant models ($P = 5.63 \times 10^{-7}$), but ranked the 75^{th} ($P = 1.07 \times 10^{-5}$) and 26^{th} ($P = 3.71 \times 10^{-6}$) in the recessive and categorical models respectively. Among the SNPs significantly associated with tenofovir clearance, SNP rs12082252 ranked 1^{st} ($P = 2.28 \times 10^{-10}$) and 2^{nd} ($P = 1.69 \times 10^{-9}$) in the recessive and categorical models respectively, but was ranked much lower in the other model specifications (346688^{th} in CoCoBOT, 1924^{th} in additive, 90918^{th} in dominant).

Results of similar analyses (top SNPs), but separately within each race/ethnicity, are provided in Appendix B Tables 17- 19. Among African Americans, 7 (2) SNPs were genome-wide significant in the recessive (categorical) model specification. SNP rs12082252 was among the significant 7 in the recessive model, and the 2 SNPs significant in the categorical model (rs4612347 and rs1632962) were also significant in the recessive model and ranked 22467^{th} and 404712^{th} respectively in the CoCoBOT model. No SNPs were genome-wide significant among Europeans, and only one SNP, rs1511185 ($P = 4.08 \times 10^{-8}$) was significant among Hispanics under the CoCoBOT model. Based on the rankings from Table 9, there appears to be a similar pattern of detection between the CoCoBOT and additive model, and between the recessive and categorical model.

Table 20 shows results of analyses of all subjects using baseline bilirubin as outcome. As expected, several polymorphisms in the UGT1A1 gene were associated with baseline plasma bilirubin; 8, 16, 8 and 9 SNPs were genome-wide significant under the CoCoBOT, additive, recessive and categorical model specifications respectively. Surprisingly, no polymorphisms were detected by the dominant model.

Table 9 TDF clearance: Rank of SNPs with the smallest p-values in combined group analysis

	CoCoBOT		**Additive**		**Dominant**		**Recessive**		**Categorical**	
SNP	rank	P	rank	P	rank	P	rank	P	rank	P
rs12387850	1	2.80e-07	1	1.26e-06	1	5.63e-07	75	1.07e-05	26	3.71e-06
rs1141034	2	2.63e-06	41	4.28e-05	6	8.05e-06	135848	1.59e-01	118	4.74e-05
rs4688755	3	4.28e-06	20	2.74e-05	67	8.39e-05	4543	3.70e-03	213	1.14e-04
rs6653311	4	4.88e-06	18	2.57e-05	50	5.67e-05	146	3.70e-05	257	1.41e-04
rs2883057	5	4.92e-06	25	2.99e-05	68	8.48e-05	4727	3.88e-03	230	1.23e-04
rs4688690	6	4.92e-06	26	2.99e-05	69	8.48e-05	4728	3.88e-03	231	1.23e-04
rs2526751	7	4.92e-06	27	2.99e-05	70	8.48e-05	4729	3.88e-03	232	1.23e-04
rs9311446	8	4.92e-06	28	2.99e-05	71	8.48e-05	4730	3.88e-03	233	1.23e-04
rs12420080	9	4.92e-06	57	5.03e-05	15	2.14e-05	716294	8.56e-01	197	1.01e-04
rs4688758	10	5.04e-06	24	2.95e-05	74	8.60e-05	4678	3.82e-03	229	1.22e-04
GA032783	23	1.68e-05	2	2.14e-06	97	1.09e-04	19	1.98e-06	4	8.75e-08
rs4891230	4024	4.53e-03	3	6.46e-06	870	9.85e-04	30	2.98e-06	19	2.30e-06
rs9950415	642	6.62e-04	4	6.76e-06	28	3.53e-05	5360	4.57e-03	107	3.98e-05
rs2478851	54	4.92e-05	5	9.58e-06	44	4.67e-05	3135	2.35e-03	127	5.41e-05
rs11165778	76	7.63e-05	6	9.59e-06	64	8.15e-05	3030	2.26e-03	110	4.22e-05
rs8001616	7408	8.60e-03	7	9.66e-06	178	1.93e-04	512	2.47e-04	45	1.42e-05
rs11723812	414	4.15e-04	8	1.19e-05	5	7.41e-06	87153	9.97e-02	105	3.87e-05
rs1375178	32	3.03e-05	9	1.27e-05	144	1.58e-04	3572	2.76e-03	148	7.03e-05
rs199529	27	2.58e-05	10	1.41e-05	18	2.84e-05	9636	9.11e-03	144	6.89e-05
rs12082252	346688	4.14e-01	1924	2.03e-03	90918	1.09e-01	1	2.28e-10	2	1.69e-09
rs337298	546883	6.54e-01	47735	5.58e-02	694376	8.31e-01	2	7.58e-10	1	7.80e-10
rs16823145	23568	2.80e-02	60	5.38e-05	4018	4.90e-03	3	3.01e-08	3	4.87e-08
rs263053	75194	8.89e-02	1293	1.28e-03	101034	1.21e-01	4	4.27e-08	5	3.02e-07
rs854207	73839	8.73e-02	2920	3.13e-03	97215	1.16e-01	5	1.51e-07	12	9.59e-07
rs966634	193015	2.29e-01	3361	3.64e-03	66317	7.95e-02	6	3.28e-07	17	1.61e-06
rs6695258	23840	2.84e-02	90	7.63e-05	4202	5.10e-03	7	3.85e-07	7	5.76e-07
rs9982266	14065	1.65e-02	576	5.28e-04	11433	1.36e-02	8	4.80e-07	13	1.07e-06
rs4391664	775385	9.29e-01	57553	6.75e-02	493459	5.90e-01	9	5.11e-07	21	2.79e-06
rs1469386	24915	2.96e-02	1572	1.63e-03	309053	3.70e-01	10	5.14e-07	23	3.11e-06
rs3815748	19	1.44e-05	11	1.43e-05	2	5.17e-06	339831	4.05e-01	88	3.11e-05
rs6878929	145	1.52e-04	86	7.16e-05	3	6.96e-06	560668	6.70e-01	54	1.60e-05
rs12929434	56	5.03e-05	44	4.44e-05	4	7.09e-06	794567	9.51e-01	79	2.45e-05
rs12440239	586	6.02e-04	1056	1.03e-03	7	1.00e-05	717434	8.58e-01	84	2.68e-05
rs7144413	45	4.18e-05	120	1.03e-04	8	1.14e-05	70748	8.01e-02	136	6.41e-05
rs7687008	292	2.84e-04	216	1.77e-04	9	1.26e-05	116180	1.35e-01	156	7.30e-05
rs1230103	48	4.53e-05	32	3.42e-05	10	1.65e-05	40370	4.42e-02	160	7.46e-05
rs7849259	416339	4.97e-01	389042	4.63e-01	375670	4.50e-01	14	6.83e-07	6	4.91e-07
rs12479145	133326	1.58e-01	82308	9.67e-02	195501	2.34e-01	46	3.32e-06	8	6.08e-07
rs17641977	29863	3.55e-02	414	3.71e-04	2089	2.45e-03	32	3.02e-06	9	6.69e-07
rs11791293	25726	3.06e-02	52	4.75e-05	784	8.86e-04	45	3.21e-06	10	6.81e-07

Creatinine clearance association analyses

Table 10 shows a pairwise correlation matrix of p-values from the combined group analyses under the five different analysis models. A similar pattern of correlation as that seen in the tenofovir pharmacokinetic analyses is observed here. In the combined group analyses, CoCoBOT is strongly correlated with the additive model($\rho = 0.769$) and the dominant model ($\rho = 0.684$), but weakly correlated with the recessive model ($\rho = 0.195$). Similar results were observed for the stratified analysis by race/ethnicity (not shown). Table 21 in Appendices shows the correlation matrix of p-values from analysis with baseline plasma bilirubin as outcome, which show a similar pattern to that seen in Table 10.

Table 10 6-month CrCl change: Correlation matrix of p-values from combined group analysis

	CoCoBOT	**Additive**	**Dominant**	**Recessive**	**Categorical**
CoCoBOT	1				
Additive	0.769	1			
Dominant	0.684	0.736	1		
Recessive	0.195	0.318	0.065	1	
Categorical	0.542	0.647	0.648	0.646	1

The 10 SNPs with the smallest p-values in the analyses of all subjects regardless of race are shown in Table 11. Of all five models, only the recessive model showed genome-wide association with three polymorphisms (rs10841159, $P = 1.82 \times 10^{-8}$; rs7297759, $P = 1.82 \times 10^{-8}$; and rs2926112, $P = 4.86 \times 10^{-8}$). Although not significant, these three SNPs were also the top three SNPs in the categorical model.

Table 11 6-month CrCl change: SNPs with the smallest p-values in combined group analysis

CoCoBOT			**Additive**			**Dominant**			**Recessive**			**Categorical**		
CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P
3	rs9820757	2.62e-06	15	rs7173993	2.28e-06	3	rs9820757	1.71e-07	12	rs10841159	1.82e-08	8	rs2926112	6.97e-08
15	rs12594783	4.67e-06	15	rs17703807	5.64e-06	1	rs1336625	9.13e-07	12	rs7297759	1.82e-08	12	rs7297759	7.74e-08
9	rs10820825	8.91e-06	15	rs12594783	5.77e-06	1	rs549319	2.76e-06	8	rs2926112	4.86e-08	12	rs10841159	9.06e-08
15	rs8033975	9.59e-06	7	rs2191496	8.32e-06	20	rs6031252	3.35e-06	13	rs7338174	1.13e-07	8	rs7003737	2.56e-07
8	rs16892482	1.03e-05	19	rs2168632	9.39e-06	3	rs9860804	4.51e-06	12	rs4274241	1.43e-07	13	rs7338174	4.86e-07
7	rs10253780	1.49e-05	6	rs6557351	1.17e-05	3	rs1112792	4.69e-06	8	rs7003737	1.89e-07	12	rs4274241	5.21e-07
7	rs2191496	1.53e-05	1	rs645945	1.26e-05	15	rs8033975	5.04e-06	11	rs2230274	2.27e-07	8	rs16902124	5.45e-07
3	rs7650374	1.64e-05	3	rs10490834	1.30e-05	17	rs8078518	8.43e-06	6	rs12333018	3.03e-07	3	rs9820757	6.25e-07
15	rs7181586	1.97e-05	1	rs1931077	1.30e-05	7	rs2191496	1.02e-05	19	rs4474816	6.81e-07	8	rs2960488	8.66e-07
1	rs1931077	2.10e-05	6	rs2185112	1.33e-05	3	rs6550032	1.40e-05	19	rs4807371	6.81e-07	11	rs2230274	1.13e-06

Table 12 shows the ranks of the unique SNPs from Table 11 under each analysis model, and their respective p-values. SNP rs10841159 ranked 25828^{th} in the CoCoBOT, 41^{st} in the additive; 12105^{th} in the dominant and 3^{rd} in the categorical models. On the other hand, rs7297759 ranked 26854^{th} in the CoCoBOT, 30^{th} in the additive, 9597^{th} in the dominant and 2^{nd} in the categorical models.

Tables 22- 24 of Appendix C show results of similar analyses (top 10 SNPs), stratified by PC-inferred

Table 12 6-month CrCl change: Rank of SNPs with the smallest p-values in combined group analysis

	CoCoBOT		Additive		Dominant		Recessive		Categorical	
SNP	rank	P	rank	P	rank	P	rank	P	rank	P
rs9820757	1	2.62e-06	36	4.77e-05	1	1.71e-07	640163	7.60e-01	8	6.25e-07
rs12594783	2	4.67e-06	3	5.77e-06	1374	1.57e-03	35	6.55e-06	24	6.75e-06
rs10820825	3	8.91e-06	12	1.57e-05	12	1.76e-05	79939	9.35e-02	121	7.78e-05
rs8033975	4	9.59e-06	53	6.50e-05	7	5.04e-06	85598	1.00e-01	57	2.99e-05
rs16892482	5	1.03e-05	52	6.45e-05	71	9.94e-05	66557	7.74e-02	402	3.39e-04
rs10253780	6	1.49e-05	28	3.93e-05	33	4.78e-05	42022	4.83e-02	212	1.65e-04
rs2191496	7	1.53e-05	4	8.32e-06	9	1.02e-05	28286	3.19e-02	68	3.94e-05
rs7650374	8	1.64e-05	11	1.51e-05	7457	8.70e-03	74	2.27e-05	67	3.90e-05
rs7181586	9	1.97e-05	50	6.40e-05	17	2.67e-05	227252	2.70e-01	190	1.48e-04
rs1931077	10	2.10e-05	9	1.30e-05	32	4.74e-05	2786	2.59e-03	92	5.87e-05
rs7173993	928	1.16e-03	1	2.28e-06	60	8.49e-05	113	4.15e-05	12	2.15e-06
rs17703807	208	2.56e-04	2	5.64e-06	67	9.25e-05	881	6.67e-04	55	2.77e-05
rs2168632	45	6.29e-05	5	9.39e-06	308	3.74e-04	358	2.17e-04	78	4.63e-05
rs6557351	35	4.85e-05	6	1.17e-05	958	1.10e-03	217	1.12e-04	65	3.88e-05
rs645945	53	7.30e-05	7	1.26e-05	164	2.08e-04	1891	1.64e-03	107	6.87e-05
rs10490834	11	2.31e-05	8	1.30e-05	16	2.48e-05	31172	3.53e-02	113	7.31e-05
rs2185112	41	5.95e-05	10	1.33e-05	1834	2.13e-03	114	4.15e-05	56	2.90e-05
rs10841159	25828	3.18e-02	41	5.53e-05	12105	1.40e-02	1	1.82e-08	3	9.06e-08
rs7297759	26854	3.30e-02	30	4.15e-05	9597	1.11e-02	2	1.82e-08	2	7.74e-08
rs2926112	832804	9.88e-01	200227	2.37e-01	581685	6.87e-01	3	4.86e-08	1	6.97e-08
rs7338174	157916	1.91e-01	20472	2.39e-02	99258	1.17e-01	4	1.13e-07	5	4.86e-07
rs4274241	28400	3.49e-02	49	6.31e-05	9042	1.05e-02	5	1.43e-07	6	5.21e-07
rs7003737	823043	9.76e-01	225009	2.66e-01	518513	6.13e-01	6	1.89e-07	4	2.56e-07
rs2230274	158597	1.92e-01	78754	9.29e-02	203969	2.40e-01	7	2.27e-07	10	1.13e-06
rs12333018	666887	7.92e-01	122506	1.44e-01	818529	9.70e-01	8	3.03e-07	11	1.18e-06
rs4474816	238439	2.87e-01	18705	2.18e-02	159513	1.87e-01	9	6.81e-07	18	4.18e-06
rs4807371	189103	2.28e-01	12143	1.40e-02	119001	1.40e-01	10	6.81e-07	17	3.96e-06
rs1336625	1143	1.42e-03	367	4.24e-04	2	9.13e-07	493673	5.87e-01	13	2.68e-06
rs549319	691	8.83e-04	400	4.61e-04	3	2.76e-06	345209	4.10e-01	37	1.16e-05
rs6031252	349	4.51e-04	211	2.66e-04	4	3.35e-06	606421	7.20e-01	40	1.36e-05
rs9860804	18	3.28e-05	321	3.77e-04	5	4.51e-06	602277	7.15e-01	47	1.78e-05
rs1112792	12	2.58e-05	665	7.41e-04	6	4.69e-06	608413	7.22e-01	25	6.83e-06
rs8078518	198	2.34e-04	54	6.52e-05	8	8.43e-06	51651	5.96e-02	80	4.70e-05
rs6550032	37	5.05e-05	410	4.72e-04	10	1.40e-05	592278	7.03e-01	88	5.54e-05
rs16902124	211388	2.55e-01	822965	9.76e-01	110320	1.30e-01	30	5.15e-06	7	5.45e-07
rs2960488	724363	8.60e-01	266864	3.15e-01	494636	5.84e-01	11	7.46e-07	9	8.66e-07

race/ethnicity. Only 2 SNPs (rs9959038, $P = 1.50 \times 10^{-8}$ and rs4934394, $P = 2.80 \times 10^{-8}$) under the recessive model specification were genome-wide significant among African Americans; these two SNPs were also among the top 3 in the categorical model, but were ranked much lower in CoCoBOT, additive and dominant models. 6 (5) SNPs were associated with 6-month CrCl change among Hispanics in the recessive (categorical) models, with a notable overlap in the top SNPs between the two models. No polymorphisms were genome-wide significant among Europeans.

Results of combined analyses using baseline bilirubin as outcome are shown in Appendix C, Table 25. Confirming that our analyses could detect true associations, multiple SNPs in $UGT1A1$ were associated with baseline bilirubin; 35, 45, 18, 15 and 28 SNPs were associated with baseline bilirubin under the CoCoBOT, additive, dominant, recessive and categorical models respectively.

Discussion

In this case study, we compared the SNP detection performance of a new method (CoCoBOT) and that of the commonly used model specifications for genetic data analysis (additive, recessive, dominant, categorical). The association between genotype and two phenotypes (tenofovir clearance and 6-month CrCl change), after adjusting for other covariates, was evaluated in a diverse group of HIV-infected subjects randomized to tenofovir-containing regimens in a prospective clinical trial.

For all combined group and stratified analyses, the pairwise correlation matrices showed strong (weak) correlation between the p-values from the CoCoBOT and additive (recessive) models. Correlations of about the same magnitude were also seen between the categorical and additive, recessive and dominant models. Although we observed this pattern of correlation, it did not always correspond to the ranks and p-values of the SNPs detected by each of the five models. For instance, in the combined group analysis with tenofovir clearance as the outcome (Table 8), the top SNP (rs12387850) in the CoCoBOT model was the same SNP in the additive and dominant models, but this similarity in ranking was not observed for the other SNPs. However, across most analyses, we saw a very similar pattern of detection between the recessive and categorical model specifications in that both models often detected the same top SNPs. Moreover, these two models tended to detect positive associations between SNPs and the two phenotypes in most of the analyses. For instance, in the tenofovir pharmacokinetic analyses, only these two models identified associations in the combined and African American groups. In the CrCl association analyses, only the recessive model detected positive associations in the combined and African American group analyses, and both the categorical and recessive models detected associations among Hispanics. These results suggest a similarity in the detection behavior of the recessive and categorical models, which could be explained partly by the possibility that most of the effect is driven by one genotype category (a/a in this case since the recessive model requires this genotype to show an effect).

No polymorphisms were significant in the CrCl association analyses under the CoCoBOT model specification. However, in the TDF pharmacokinetic analyses, CoCoBOT detected a positive association with rs1511185 in the CTNNA2 gene (chromosome 2) among Hispanics. Although this is not one of the SNPs known to be linked to tenofovir-related phenotypes, it showed the ability to detect positive associations with CoCoBOT. In addition, there was some overlap in the top SNPs identified by the CoCoBOT and additive models, especially with the known *UGT1A1* polymorphisms in the baseline bilirubin analyses (Tables 20 and 25).

Among the SNPs that were positively associated with tenofovir clearance in the combined group analyses under the recessive model, rs12082252 is located in the TIPRL gene, rs337298 is located in the

WDR47 gene and rs16823145 is located in the C1orf21 gene. None of these genes have been linked to tenofovir-related phenotypes, and their significance to this study is not clear. Of all the top SNPs shown in Table 8 and 11, none were SNPs that have been identified to be associated with tenofovir pharmacokinetics and kidney-related phenotypes respectively.

Although the combined group analyses had over 500 subjects, the stratified analyses per race/ethnicity group had few individuals, possibly resulting in low power to detect SNPs across the five analysis models. Another limitation of this study specific to CoCoBOT is that presently, it cannot handle testing for interaction effects between genotype and other covariates as may be desired in some situations. As a result, we could not for instance, test for interactions with genotype and treatment arm (TDF or abacavir) as was done in Chapter II with the additive model. Thus, extension of CoCoBOT to handle interactions with the ordinal predictor is future work.

In summary, we identified some significant associations with TDF clearance and CrCl, especially using the recessive and categorical models. However, since most of the polymorphisms were not any of the ones that have been reported in previous studies to be associated with tenofovir or CrCl-related phenotypes, it is unclear if they are new hits, or whether the identified effects were driven by some unknown factor. Candidate gene studies could shed more light into these findings, and further research in different study populations receiving other tenofovir-containing regimens is necessary.

Finally, CoCoBOT was seen to have the ability to detect SNPs, just like the other model specifications, especially with known *UGT1A1* SNPs and baseline plasma bilirubin concentration. Thus, CoCoBOT could be an option to consider in genetic data analyses because of the fewer model assumptions it makes when a predictor is ordinal. However, based on the simulation results, using CoCoBOT did not seem to result in much gain in power to detect genetic associations; except under the truth, the additive and categorical models seemed to perform more robustly in most data generation scenarios, including the non-linear settings. Evaluating the performance of CoCoBOT in candidate gene studies and in cases where genotype is interacted with other covariates may yield different results, but based on the findings from this study, the additive or categorical models are adequately powerful to detect associations.

CHAPTER V

Conclusions

In this project, we introduced a new method (CoCoBOT) to detect SNPs in a genome-wide association study. Using simulations, we evaluated and discussed the statistical performance of CoCoBOT and compared this to the performance of additive, dominant, recessive and categorical models in a setting where the outcome is continuous and the main predictor, genotype, is ordered categorical. The motivation was that unlike the other models, CoCoBOT allows a monotone relationship between the outcome and the ordinal predictor. The comparison of the power of these methods to detect genetic associations indicated that CoCoBOT does not appear to be the best choice for genetic association studies. Like the other models, CoCoBOT had high power under higher minor allele frequencies (MAFs) and struggled under low MAFs, especially when the data are generated under the recessive model. This is believed to be due to the fact that at such low frequencies, the genotype effect is mainly due to the minor allele, a and the probability of observing the a/a genotype category is very low. In addition, the low power under the recessive model is believed to be due to the fact that the main effect is driven by only one category (a/a), and the probability of observing this category is very low.

Although in some instances CoCoBOT performed better than the additive model, their power to detect associations did not differ by large magnitudes, resulting in no unequivocal winner between the two models in settings other than the truth. Except under the properly specified models, the categorical model did not lag behind in comparison to the CoCoBOT and additive models; for instance, at MAF = 0.5 under the dominant and recessive data generation models, the categorical model outperformed both CoCoBOT and the additive model. Furthermore, at MAFs of 0.3 and 0.1, the categorical model was not the worst of the other of the models that were not the truth. This indicates that in situations where the data generation model is unknown, the categorical model may be a robust choice.

Based on the case study using ACTG Protocol A5202 data, CoCoBOT had the ability to detect SNPs. Although the recessive and categorical models detected most positive associations, none of them were known associations with tenofovir clearance or changes in creatinine clearance. This leaves a new area for future investigations. Candidate SNPs/genes analyses may reveal more about the detection performance of CoCoBOT and the other four models with polymorphisms that have been reported in literature to be associated with tenofovir clearance and kidney-related phenotypes.

A strength of the analyses we performed was that it involved individuals from a diverse population, and the combined analyses allowed the evaluation of over 500 samples. However, it is possible that small sample

sizes may have contributed to low detection power in the stratified analyses by race/ethnicity. Therefore, stratified analyses with larger samples could yield higher detection power.

In Chapter II, where we conducted analyses using an additive model with 6-month CrCl change as the outcome, we were able to include the interaction of genotype and treatment group (tenofovir or abacavir). However, in the case study in Chapter IV, we were unable to test for this interaction as desired due to CoCoBOT's present inability to handle interactions. Therefore, improvements of the method to allow interactions with ordinal predictors need to be made to allow the evaluation of a wide array of research questions.

In conclusion, based on the simulations carried out, the additive, categorical and CoCoBOT models, in that order, would be recommended for analysis of genetic associations. This is because it the model that real data follow is typically not known, and when recessive model is true, the dominant model performs worse than the other models and when the dominant model is true, the recessive models performs worse than the other models. However, it is important to note that not much detection power is gained in using CoCoBOT when compared to the additive or categorical models in most of the settings we evaluated This is seen both in the results from the simulations and case study with real data. Thus, CoCoBOT can be suitable for exploratory purposes on detection performance of SNPs; beyond that the currently avaialble methods are robust when chosen appropriately.

Appendices

A Tenofovir pharmacokinetic (PK) and creatinine clearance association results from Chapter II

Table 13 shows the 20 SNPs with lowest p-values in each analysis with tenofovir clearance as the outcome.

Table 14 shows the 20 SNPs with lowest p-values in each analysis with *time-dependent CrCl change* as the outcome.

Table 15 shows the 20 SNPs with lowest p-values in each sensitivity analysis with 6-month CrCl change as the outcome.

Table 13 Tenofovir Pharmacokinetic (PK) Association Results (20 SNPs with lowest p-values in each analysis)

Combined Group			African Americans			Europeans			Hispanics		
SNP	CHR	P	SNP	CHR	P	SNP	CHR	P	SNP	CHR	P
Genome-wide SNPs											
GA032783	17	2.1e-06	rs7645493	3	2.5e-06	rs6575267	14	2.1e-07	rs9648724	7	1.1e-06
rs4891230	18	6.5e-06	rs8040826	15	3.5e-06	rs7160376	14	2.3e-06	rs2398827	9	3.1e-06
rs9950415	18	6.8e-06	rs2570249	15	4.8e-06	rs4401457	4	5.2e-06	rs4282783	1	6.5e-06
rs2478851	9	9.6e-06	rs2570218	15	4.8e-06	rs10013079	4	5.4e-06	rs10912094	1	9.3e-06
rs11165778	1	9.6e-06	rs2554341	15	6.1e-06	rs1317816	6	5.4e-06	rs1732103	4	1.3e-05
rs8001616	13	9.7e-06	rs2554321	15	6.1e-06	rs4712976	6	7.7e-06	rs1250105	4	1.3e-05
rs11723812	4	1.2e-05	rs4309482	18	8.8e-06	rs718763	2	9.6e-06	rs4357909	15	1.5e-05
rs1375178	2	1.3e-05	rs10823406	10	9.3e-06	rs609949	18	1.3e-05	rs4743894	9	1.7e-05
rs199529	17	1.4e-05	rs4858429	3	9.7e-06	rs7749149	6	1.3e-05	rs10985148	9	1.7e-05
rs3815748	19	1.4e-05	rs11906888	20	1.0e-05	rs567338	18	1.6e-05	rs4710994	6	2.1e-05
rs9854936	3	1.8e-05	rs2200488	2	1.2e-05	rs1408271	6	1.7e-05	rs6137387	20	2.1e-05
rs6999964	8	1.9e-05	rs6133987	20	1.4e-05	rs2720823	8	1.9e-05	rs12718174	7	2.3e-05
rs1548896	2	2.1e-05	rs369427	1	1.9e-05	rs7653963	4	2.0e-05	rs10227315	7	2.3e-05
rs4074408	18	2.2e-05	rs11177175	12	1.9e-05	rs3778272	6	2.0e-05	rs10040797	5	2.6e-05
rs17199679	9	2.4e-05	rs3762872	16	2.1e-05	rs1892248	6	2.0e-05	rs12578725	12	2.7e-05
rs6429178	1	2.5e-05	rs2289430	16	2.1e-05	rs4712970	6	2.0e-05	rs7548935	1	2.8e-05
rs4688755	3	2.7e-05	rs2277968	19	2.3e-05	rs1317510	6	2.0e-05	rs12080794	1	2.9e-05
rs1382101	8	2.8e-05	rs9865933	3	2.7e-05	rs9759759	4	2.0e-05	rs8008246	14	2.9e-05
rs247938	5	2.9e-05	rs126917	20	2.8e-05	rs10031444	4	2.2e-05	rs10993725	9	3.0e-05
rs4688758	3	3.0e-05	rs12211633	6	2.8e-05	rs10000845	4	2.2e-05	rs876670	10	3.1e-05
30 Candidate SNPs											
rs456374	9	0.002268	rs465793	9	0.00025	rs1729741	13	0.02182	rs17268122	13	0.000436
rs465793	9	0.003149	rs456374	9	0.00311	rs17268170	13	0.0253	rs17268163	13	0.000580
rs16921966	9	0.006022	rs9561811	13	0.00930	rs17268129	13	0.03852	rs2159359	17	0.000653
rs1189462	13	0.007878	rs2274410	13	0.00999	rs873705	13	0.04253	rs2041296	17	0.000929
rs3847258	9	0.009197	rs6949448	7	0.01043	rs1751027	13	0.04335	rs12864844	13	0.001003
rs9561765	13	0.01058	rs10276036	7	0.01186	rs11231302	11	0.04396	rs17189446	13	0.001342
rs1189461	13	0.01405	rs3818494	13	0.01207	rs2159359	17	0.05364	rs12864049	13	0.001807
rs1751037	13	0.01405	rs12704364	7	0.01235	rs11591185	1	0.06099	rs1479389	13	0.003163
rs7330330	13	0.01616	rs6961665	7	0.01235	rs4773840	13	0.06527	rs8075231	17	0.003591
rs403860	9	0.01635	rs2274408	13	0.01579	rs1611822	13	0.06789	rs17189376	13	0.003705
rs4148540	13	0.02114	rs1202170	7	0.01682	rs1564351	13	0.06885	rs12584534	13	0.00404
rs7331488	13	0.02114	rs17189481	13	0.01761	rs4148487	13	0.06947	rs4771912	13	0.004216
rs1189464	13	0.02581	rs2235046	7	0.01892	rs1678341	13	0.07359	rs4773843	13	0.01593
rs1189466	13	0.0275	rs2235033	7	0.01925	rs6492768	13	0.07713	rs17189299	13	0.01671
rs1678387	13	0.0275	rs1202169	7	0.02112	rs12429339	13	0.09089	rs9524822	13	0.01918
rs1678409	13	0.02843	rs10274587	7	0.02313	rs1678384	13	0.09237	rs1611822	13	0.01979
rs1678365	13	0.02887	rs2235040	7	0.02313	rs10161985	13	0.09639	rs9524849	13	0.02633
rs4148430	13	0.02938	rs7981095	13	0.0232	rs2185631	6	0.09751	rs1048020	9	0.02969
rs1729745	13	0.02955	rs4148747	7	0.02458	rs9349256	6	0.09956	rs1729764	13	0.02996
rs2766475	13	0.02987	rs1202172	7	0.02498	rs9394952	6	0.09956	rs3765535	13	0.02996
594 Candidate SNPs											
rs456374	9	0.002268	rs465793	9	0.00025	rs1729741	13	0.02182	rs17268122	13	0.000436
rs465793	9	0.003149	rs456374	9	0.00311	rs17268170	13	0.0253	rs17268163	13	0.000580
rs16921966	9	0.006022	rs9561811	13	0.00930	rs17268129	13	0.03852	rs2159359	17	0.000653
rs1189462	13	0.007878	rs2274410	13	0.00999	rs873705	13	0.04253	rs2041296	17	0.000929
rs3847258	9	0.009197	rs6949448	7	0.01043	rs1751027	13	0.04335	rs12864844	13	0.001003
rs9561765	13	0.01058	rs10276036	7	0.01186	rs11231302	11	0.04396	rs17189446	13	0.001342
rs1189461	13	0.01405	rs3818494	13	0.01207	rs2159359	17	0.05364	rs12864049	13	0.001807
rs1751037	13	0.01405	rs12704364	7	0.01235	rs11591185	1	0.06099	rs1479389	13	0.003163
rs7330330	13	0.01616	rs6961665	7	0.01235	rs4773840	13	0.06527	rs8075231	17	0.003591
rs403860	9	0.01635	rs2274408	13	0.01579	rs1611822	13	0.06789	rs17189376	13	0.003705
rs4148540	13	0.02114	rs1202170	7	0.01682	rs1564351	13	0.06885	rs12584534	13	0.00404
rs7331488	13	0.02114	rs17189481	13	0.01761	rs4148487	13	0.06947	rs4771912	13	0.004216
rs1189464	13	0.02581	rs2235046	7	0.01892	rs1678341	13	0.07359	rs4773843	13	0.01593
rs1189466	13	0.0275	rs2235033	7	0.01925	rs6492768	13	0.07713	rs17189299	13	0.01671
rs1678387	13	0.0275	rs1202169	7	0.02112	rs12429339	13	0.09089	rs9524822	13	0.01918
rs1678409	13	0.02843	rs10274587	7	0.02313	rs1678384	13	0.09237	rs1611822	13	0.01979
rs1678365	13	0.02887	rs2235040	7	0.02313	rs10161985	13	0.09639	rs9524849	13	0.02633
rs4148430	13	0.02938	rs7981095	13	0.0232	rs2185631	6	0.09751	rs1048020	9	0.02969
rs1729745	13	0.02955	rs4148747	7	0.02458	rs9349256	6	0.09956	rs1729764	13	0.02996
rs2766475	13	0.02987	rs1202172	7	0.02498	rs9394952	6	0.09956	rs3765535	13	0.02996

Significance threshold was: 5×10^{-8} for genome-wide SNPs, 0.002 for the subset of 30 SNPs, 8.4×10^{-5} for the subset of 594 SNPs.

Table 14 Creatinine Clearance Association Results (20 SNPs with lowest p-values in each analysis)

Combined Group			African Americans			Europeans			Hispanics		
SNP	CHR	P	SNP	CHR	P	SNP	CHR	P	SNP	CHR	P
Genome-wide SNPs											
rs4435343	18	1.24e-06	rs10827791	10	1.79e-06	rs11114652	12	2.91e-06	rs17062791	8	1.77e-07
rs6533668	4	9.28e-06	rs11106805	12	3.38e-06	rs4962347	10	4.35e-06	rs4936767	11	6.22e-07
rs2352185	4	9.92e-06	rs10786737	10	3.60e-06	rs17804080	12	4.79e-06	rs7109445	11	7.08e-07
rs2011706	21	1.12e-05	rs6863176	5	3.76e-06	rs16918212	12	6.76e-06	rs4936770	11	8.86e-07
rs11082117	18	1.29e-05	rs7996510	13	5.70e-06	rs7585799	2	7.35e-06	rs2226150	6	1.95e-06
rs12953983	18	1.48e-05	rs2571468	4	5.95e-06	rs10399860	1	8.79e-06	rs1592216	6	2.50e-06
rs9323158	14	1.66e-05	rs10827797	10	6.95e-06	rs2129530	11	9.30e-06	rs2883821	1	3.07e-06
rs12597817	16	1.89e-05	rs10521695	23	8.27e-06	rs7903887	10	9.85e-06	rs6677658	1	3.33e-06
rs2407940	8	1.89e-05	rs534613	23	1.41e-05	rs7214860	17	1.09e-05	rs7688805	4	3.73e-06
rs4521822	9	2.00e-05	rs1450657	5	1.51e-05	rs6548164	2	1.17e-05	rs505601	18	4.51e-06
rs656682	10	2.00e-05	rs11953822	5	1.87e-05	rs9576310	13	1.43e-05	rs1935737	6	4.62e-06
rs1364182	16	2.16e-05	rs4242134	5	1.99e-05	rs4514905	2	1.47e-05	rs1342634	6	4.73e-06
rs10853461	18	2.17e-05	rs2249751	6	2.14e-05	rs3812842	13	1.48e-05	rs7064462	23	5.26e-06
rs1886970	13	2.28e-05	rs13711	11	2.21e-05	rs11613504	12	1.88e-05	rs7562658	2	5.65e-06
rs1751036	13	2.41e-05	rs17043253	3	2.26e-05	rs2025405	13	1.88e-05	rs6902991	6	6.09e-06
rs17878498	3	2.41e-05	rs12422149	11	2.31e-05	rs13311400	7	2.09e-05	rs7923837	10	6.71e-06
rs4825731	23	2.43e-05	rs4303	17	2.39e-05	rs1696320	12	2.09e-05	rs4709724	6	6.89e-06
rs11621998	14	2.52e-05	rs10929525	2	2.51e-05	rs2079778	17	2.16e-05	rs7941144	11	7.75e-06
rs4243086	15	2.55e-05	rs8004116	14	2.53e-05	rs1924296	13	2.28e-05	rs12592731	15	7.91e-06
rs1900100	14	2.59e-05	rs11693395	2	2.56e-05	rs10741552	11	2.28e-05	rs36111427	6	8.96e-06
212 Candidate SNPs											
rs2082424	19	0.0054	rs3127573	6	0.0018	rs2412971	22	0.0087	rs6036478	20	0.0039
rs7246178	19	0.0058	rs1556751	9	0.0129	rs1260326	2	0.0095	rs6999484	8	0.0086
rs9275596	6	0.0223	rs10941692	5	0.0448	rs2764267	6	0.0311	rs6048952	20	0.0195
rs17272197	19	0.0241	rs923068	18	0.0482	rs1775644	6	0.0360	rs4346460	20	0.0212
rs653178	12	0.0291	rs12520150	5	0.0664	rs948494	11	0.0414	rs7805747	7	0.0224
rs7805747	7	0.0405	rs11871637	17	0.0821	rs12537	22	0.0435	rs2728108	4	0.0231
rs10941692	5	0.0449	rs6465825	7	0.0822	rs16946160	13	0.0469	rs3810575	20	0.0240
rs6677604	1	0.0467	rs1133029	20	0.0835	rs11227281	11	0.0500	rs6677604	1	0.0291
rs12520150	5	0.0478	rs12514615	5	0.0889	rs6420094	5	0.0922	rs10794720	10	0.0417
rs10774021	12	0.0488	rs12522822	5	0.0959	rs9473932	6	0.1021	rs1705699	8	0.0480
rs660895	6	0.0513	rs16902083	5	0.0959	rs2057291	20	0.1035	rs35610040	20	0.0598
rs2239785	22	0.0540	rs4566805	5	0.0959	rs2518322	6	0.1058	rs17751897	20	0.0671
rs9473932	6	0.0687	rs12515820	5	0.1005	rs17272197	19	0.1206	rs3787498	20	0.0725
rs3115573	6	0.0799	rs3813227	2	0.1174	rs2082424	19	0.1216	rs3827142	20	0.0736
rs3127573	6	0.0891	rs10184268	2	0.1178	rs4812042	20	0.1303	rs3787499	20	0.0763
rs1556751	9	0.0924	rs1909937	5	0.1294	rs660895	6	0.1314	rs1719250	15	0.0873
rs1133029	20	0.1085	rs9473932	6	0.1337	rs7246178	19	0.1345	rs3827143	20	0.0906
rs7105665	11	0.1217	rs3115573	6	0.1658	rs6026576	20	0.1355	rs12625716	20	0.0939
rs11871637	17	0.1262	rs1392970	5	0.1865	rs2239785	22	0.1393	rs653178	12	0.0942
rs2277311	11	0.1309	rs6420094	5	0.1889	rs10941692	5	0.1521	rs4664308	2	0.1187

Significance threshold was: 5×10^{-8} for genome-wide SNPs, 0.0002 for the subset of 212 SNPs.

Table 15 Sensitivity Analysis Results for CrCl (20 SNPs with lowest p-values in each analysis)

Combined Group			African Americans			Europeans			Hispanics			Meta-analysis		
SNP	CHR	P	SNP	CHR	P	SNP	CHR	P	SNP	CHR	P	SNP	CHR	P
Genome-wide SNPs														
rs1397050	11	4.58e-06	rs2480181	6	4.49e-06	rs731580	3	1.03e-06	rs6448638	4	6.79e-07	rs906956	11	4.18e-06
rs121882	5	4.70e-06	rs11106805	12	6.76e-06	rs1733826	7	3.56e-06	rs4435343	18	8.15e-07	rs121882	5	5.62e-06
rs906956	11	6.91e-06	rs12230258	12	9.21e-06	rs2112424	5	3.56e-06	rs1966678	4	1.13e-06	rs11227719	11	5.71e-06
rs13071033	3	7.22e-06	rs309543	1	1.27e-05	rs2438466	5	4.07e-06	rs2226150	6	1.30e-06	GA021270	3	6.30e-06
rs10510580	3	7.22e-06	rs555740	12	1.41e-05	rs16993962	20	7.70e-06	rs16893073	5	1.31e-06	rs1397050	11	6.68e-06
rs4435343	18	1.21e-05	rs11926663	3	2.11e-05	rs2664570	20	8.70e-06	rs13292946	9	1.86e-06	rs17150286	11	8.77e-06
rs193890	7	1.63e-05	rs16864810	2	2.13e-05	rs16994003	20	1.12e-05	rs16867887	4	2.45e-06	rs1842674	11	9.71e-06
rs11227719	11	2.07e-05	rs9373685	6	2.25e-05	rs4809716	20	1.14e-05	rs11103684	9	2.48e-06	GA017257	20	9.97e-06
rs679309	11	2.31e-05	rs4543114	4	2.28e-05	rs729664	20	1.14e-05	rs11834	9	2.48e-06	GA035558	20	1.00e-05
rs12498133	3	2.32e-05	rs1988833	8	2.67e-05	rs11032561	11	1.16e-05	rs2604271	9	2.70e-06	rs16993997	20	1.05e-05
rs193894	7	2.40e-05	rs6862529	5	2.70e-05	rs3757527	7	1.22e-05	rs1592216	6	2.89e-06	rs11227874	11	1.27e-05
rs3117092	1	2.63e-05	rs7088661	10	2.80e-05	rs2294910	20	1.24e-05	rs1935737	6	2.95e-06	rs17286324	3	1.44e-05
rs3117091	1	2.70e-05	rs7651518	3	2.90e-05	rs655754	3	1.86e-05	rs7562658	2	3.08e-06	rs4809719	20	1.49e-05
rs1789041	18	2.72e-05	rs12086858	1	2.93e-05	rs2058298	16	1.89e-05	rs1342634	6	3.59e-06	rs11590226	1	1.61e-05
rs1524930	21	2.85e-05	rs6768930	3	3.07e-05	rs830995	2	2.01e-05	rs625001	18	3.80e-06	rs10896271	11	1.67e-05
rs1842674	11	2.96e-05	rs3127573	6	3.28e-05	rs1003615	16	2.07e-05	rs11103683	9	3.99e-06	rs2833143	21	1.73e-05
rs957215	3	2.96e-05	rs34740624	11	3.43e-05	rs1347382	8	2.08e-05	rs6902991	6	4.00e-06	rs12365860	11	2.09e-05
rs3731711	2	3.03e-05	rs16883776	5	3.50e-05	rs1397050	11	2.42e-05	rs9361224	6	4.69e-06	rs7118155	11	2.24e-05
rs12497411	3	3.09e-05	rs10789496	1	4.03e-05	rs906956	11	2.42e-05	rs904200	16	6.67e-06	rs13071033	3	2.41e-05
rs1862409	5	3.52e-05	rs11713641	3	4.13e-05	rs1812591	16	2.49e-05	rs11981075	7	6.76e-06	rs10510580	3	2.41e-05
212 Candidate SNPs														
rs17751897	20	0.00453	rs3127573 c	6	3.28e-05	rs1260326	2	0.01477	rs6036478	20	0.003525	rs10774021	12	0.01532
rs6048952	20	0.00704	rs1556751	9	0.01223	rs10774021	12	0.01952	rs3827142	20	0.003709	rs1260326	2	0.03362
rs2239785	22	0.007497	rs923068	18	0.02742	rs1883414	6	0.03378	rs35610040	20	0.004377	rs11142	1	0.08887
rs12625716	20	0.01177	rs16946160	13	0.03828	rs7422339	2	0.04907	rs13043610	20	0.007981	rs2564002	2	0.0901
rs4346460	20	0.01471	rs17751897	20	0.04128	rs3115573	6	0.04985	rs3787499	20	0.008277	rs17751897	20	0.1003
rs3810575	20	0.01835	rs12625716	20	0.06595	rs2412971	22	0.05236	rs3787498	20	0.008472	rs1158167	20	0.1136
rs1158167	20	0.0272	rs6048952	20	0.07615	rs2239785	22	0.06195	rs3810575	20	0.01057	rs4346460	20	0.1175
rs10774021	12	0.02726	rs9661614	1	0.09001	rs10745354	1	0.0762	rs12523157	5	0.01085	rs12625716	20	0.1222
rs13043610	20	0.03573	rs2564002	2	0.09124	rs4970760	1	0.07879	rs6048952	20	0.0115	rs347685	3	0.1263
rs3127573	6	0.03634	rs3115573	6	0.09299	rs10745352	1	0.08168	rs12515179	5	0.01183	rs4970759	1	0.131
rs6036478	20	0.05022	rs12136063	1	0.097	rs10858085	1	0.08168	rs3827143	20	0.01273	rs6048952	20	0.1465
rs13037490	20	0.07049	rs4928134	3	0.09921	rs10858086	1	0.08168	rs1775644	6	0.01296	rs9310709	3	0.1545
rs1556751	9	0.09159	rs11924318	3	0.1005	rs10858092	1	0.08168	rs17751897	20	0.01312	rs3810575	20	0.1591
rs2722583	2	0.09908	rs12654812	5	0.1033	rs1880670	1	0.08168	rs4346460	20	0.01741	rs7246178	19	0.1764
rs11142	1	0.1046	rs1394125	15	0.1035	rs3768497	1	0.08168	rs12625716	20	0.0229	rs13043610	20	0.1787
rs1260326	2	0.1103	rs12145677	1	0.1036	rs3853501	1	0.08168	rs10794720	10	0.02499	rs12073497	1	0.1797
rs12523157	5	0.1169	rs4849121	2	0.1131	rs3879450	1	0.08168	rs13037490	20	0.02687	rs4970760	1	0.1824
rs2564002	2	0.1304	rs6879012	5	0.1148	rs443345	1	0.08168	rs2239785	22	0.03288	rs10858086	1	0.1831
rs16946160	13	0.1328	rs4849179	2	0.1251	rs444387	1	0.08168	rs12516998	5	0.03385	rs3768497	1	0.1831
rs12073497	1	0.1332	rs1158167	20	0.1271	rs4603158	1	0.08168	rs7711446	5	0.03385	rs1133029	20	0.1839

Significance threshold was: 5×10^{-8} for genome-wide SNPs, 0.0002 for the subset of 212 SNPs. rs3127573 on chromosome 6 of the SLC22A2 gene was significant in the African American group analysis of candidate SNPs.

B Chapter IV (Tenofovir pharmacokinetic association analyses)

Table 16 shows the pairwise correlation matrix of p-values from analysis of all subjects with baseline bilirubin as outcome

Table 16 Baseline bilirubin: Correlation matrix of p-values from combined group analysis

	CoCoBOT	**Additive**	**Dominant**	**Recessive**	**Categorical**
CoCoBOT	1				
Additive	0.818	1			
Dominant	0.724	0.738	1		
Recessive	0.206	0.323	0.069	1	
Categorical	0.569	0.65	0.65	0.647	1

Tables 17- 19 show the top 10 SNPs with the smallest p-values within each race/ethnicity group under each of the five analysis models with tenofovir clearance as outcome

Table 17 TDF clearance: SNPs with the smallest p-values among African Americans

CoCoBOT			**Additive**			**Dominant**			**Recessive**			**Categorical**		
CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P
2	rs3850354	6.64e-07	3	rs7645493	2.50e-06	8	rs2382993	2.98e-06	1	rs12082252	3.08e-08	8	rs4612347	3.42e-08
15	rs8040826	1.59e-06	15	rs8040826	3.48e-06	3	rs6781193	6.37e-06	9	rs4554572	3.08e-08	6	rs1632962	4.98e-08
17	rs17627030	1.65e-06	15	rs2570249	4.78e-06	2	rs266216	6.65e-06	6	rs1632962	3.21e-08	6	rs1632964	6.82e-08
11	rs3108805	2.76e-06	15	rs2570218	4.78e-06	2	rs6531176	6.79e-06	8	rs4612347	3.21e-08	1	rs12082252	1.97e-07
9	rs2382407	3.22e-06	15	rs2554341	6.13e-06	15	rs2570249	7.67e-06	6	rs1632964	3.59e-08	1	rs337298	2.02e-07
3	rs13325221	4.68e-06	15	rs2554321	6.13e-06	15	rs2570218	7.67e-06	8	rs3802143	5.91e-08	9	rs4554572	2.12e-07
11	rs2702676	5.33e-06	18	rs4309482	8.76e-06	15	rs8040826	8.13e-06	8	rs3898300	5.91e-08	8	rs3802143	2.29e-07
1	rs6588444	6.06e-06	10	rs10823406	9.32e-06	11	rs3108805	1.12e-05	1	rs337298	9.50e-08	8	rs3898300	2.29e-07
9	rs1408314	7.59e-06	20	rs11906888	1.01e-05	15	rs2554341	1.15e-05	10	rs11598684	4.75e-07	20	rs1543474	5.09e-07
14	rs7159296	1.01e-05	2	rs2200488	1.22e-05	15	rs2554321	1.15e-05	1	rs369427	5.70e-07	12	rs2250499	2.01e-06

Table 18 TDF clearance: SNPs with the smallest p-values among Europeans

CoCoBOT			**Additive**			**Dominant**			**Recessive**			**Categorical**		
CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P
14	rs6575267	8.09e-08	14	rs6575267	2.09e-07	14	rs7160376	4.71e-07	2	rs768811	6.64e-07	14	rs6575267	1.29e-06
14	rs7160376	1.99e-07	14	rs7160376	2.30e-06	14	rs6575267	1.11e-06	2	rs1053895	1.05e-06	14	rs7160376	2.62e-06
18	rs609949	1.41e-06	4	rs4401457	5.17e-06	6	rs1317816	1.72e-06	1	rs17125608	1.34e-06	2	rs768811	2.70e-06
10	rs7096455	1.46e-06	4	rs10013079	5.39e-06	6	rs4712976	3.93e-06	2	rs7422272	1.55e-06	2	rs1053895	3.97e-06
10	rs4750961	2.73e-06	6	rs1317816	5.41e-06	6	rs7749149	5.65e-06	5	rs7731904	1.55e-06	1	rs17125608	8.66e-06
6	rs1317816	3.68e-06	6	rs4712976	7.69e-06	6	rs3778272	6.75e-06	21	rs17766637	1.88e-06	2	rs7422272	8.94e-06
2	rs718763	4.51e-06	2	rs718763	9.55e-06	6	rs1892248	6.75e-06	2	rs838731	2.85e-06	1	rs7530493	9.52e-06
6	rs4712976	5.14e-06	18	rs609949	1.27e-05	6	rs4712970	6.75e-06	2	rs838732	2.96e-06	4	rs10020303	9.66e-06
17	rs9914092	6.07e-06	6	rs7749149	1.32e-05	6	rs1317510	6.75e-06	2	rs838715	3.47e-06	6	rs1317816	9.75e-06
6	rs7749149	6.28e-06	18	rs567338	1.55e-05	6	rs1141034	8.25e-06	1	rs1415105	3.70e-06	5	rs7731904	9.88e-06

Table 19 TDF clearance: SNPs with the smallest p-values among Hispanics

CoCoBOT			Additive			Dominant			Recessive			Categorical		
CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P
2	rs1511185	4.08e-08	7	rs9648724	1.09e-06	9	rs2398827	3.40e-06	11	rs11037953	6.38e-06	7	rs9648724	7.32e-06
9	rs2398827	7.49e-08	9	rs2398827	3.07e-06	7	rs9648724	3.77e-06	7	rs12718174	1.54e-05	20	rs12626158	1.02e-05
9	rs4743894	3.20e-07	1	rs10912094	9.32e-06	8	rs2114018	5.89e-06	7	rs10227315	1.54e-05	9	rs2398827	1.20e-05
7	rs9648724	4.12e-07	4	rs1732103	1.30e-05	8	rs16930786	5.89e-06	5	rs10454913	1.72e-05	22	rs8135417	1.31e-05
1	rs6677872	5.55e-07	4	rs1250105	1.32e-05	8	rs6531002	6.19e-06	7	rs2116020	2.02e-05	6	rs9386463	1.78e-05
14	rs8008246	5.65e-07	15	rs4357909	1.45e-05	20	rs12626158	6.22e-06	15	rs11259921	2.47e-05	7	rs12718174	1.86e-05
6	rs510667	7.50e-07	9	rs4743894	1.70e-05	8	rs7828391	6.36e-06	15	rs4357909	2.52e-05	7	rs10227315	1.86e-05
12	rs7957621	1.62e-06	9	rs10985148	1.71e-05	10	rs876670	6.61e-06	14	rs9323989	2.52e-05	22	rs5997893	1.92e-05
1	rs12080794	1.76e-06	6	rs4710994	2.06e-05	20	rs6137387	7.11e-06	14	rs7148578	2.52e-05	1	rs7523927	3.04e-05
6	rs4710994	1.81e-06	20	rs6137387	2.10e-05	14	rs8008246	7.23e-06	15	rs11633383	2.58e-05	1	rs7517729	3.04e-05

Table 20 shows the top 20 SNPs with the smallest p-values in the analysis of all subjects under each of the five analysis models with baseline plasma bilirubin as outcome

Table 20 Baseline bilirubin: SNPs with the smallest p-values in combined group analysis

CoCoBOT			Additive			Dominant			Recessive			Categorical		
CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P
2	rs887829	7.08e-10	2	rs887829	2.24e-11	2	rs3755319	1.03e-07	2	rs887829	4.49e-12	2	rs887829	1.74e-12
2	rs4148325	2.54e-09	2	rs4148325	7.30e-11	19	rs4239638	3.74e-07	2	rs4148325	8.81e-12	2	rs4148325	4.60e-12
2	rs6742078	4.87e-09	2	rs6742078	1.32e-10	19	rs7257832	4.52e-07	2	rs6742078	5.13e-11	2	rs6742078	1.69e-11
2	rs4148324	1.04e-08	2	rs4148324	3.11e-10	2	rs4663333	4.68e-07	2	rs929596	1.02e-10	2	rs929596	2.48e-11
2	rs10179091	1.63e-08	2	rs929596	3.39e-10	2	rs4663967	5.17e-07	2	rs4148324	1.05e-10	2	rs4148324	4.39e-11
2	rs3771341	3.07e-08	2	rs3771341	3.89e-10	2	rs4399719	8.00e-07	2	rs3771341	3.36e-10	2	rs3771341	6.07e-11
2	rs929596	3.63e-08	2	rs10179091	3.15e-09	2	rs4124874	1.03e-06	2	rs17862875	6.46e-09	2	rs17862875	2.92e-09
2	rs3755319	4.34e-08	2	rs17862875	1.56e-08	2	rs4663965	1.48e-06	2	rs10179091	2.48e-08	2	rs10179091	1.01e-08
2	rs4148326	6.53e-08	2	rs2221198	2.04e-08	2	rs6431628	2.17e-06	2	rs4148326	6.00e-08	2	rs4148326	5.80e-08
2	rs2221198	1.34e-07	2	rs4663969	2.24e-08	11	rs1560994	2.50e-06	2	rs2221198	1.86e-07	2	rs2221198	6.47e-08
2	rs4663969	1.93e-07	2	rs3755319	2.38e-08	2	rs17862866	2.77e-06	2	rs4663969	2.68e-07	2	rs3755319	7.48e-08
2	rs4663967	2.43e-07	2	rs4148326	2.73e-08	2	rs3806597	2.96e-06	2	rs16862202	2.69e-07	2	rs7604115	7.53e-08
2	rs4663333	2.81e-07	2	rs7604115	2.94e-08	2	rs2008595	3.90e-06	2	rs7556676	4.01e-07	2	rs4663969	7.94e-08
2	rs7556676	2.84e-07	2	rs7556676	3.05e-08	2	rs4294999	3.91e-06	2	rs7604115	4.46e-07	2	rs7556676	1.15e-07
2	rs871514	4.07e-07	2	rs871514	4.00e-08	10	rs7915217	4.02e-06	19	rs8111761	2.67e-06	2	rs4663967	2.53e-07
2	rs4294999	4.80e-07	2	rs4663967	5.55e-08	2	rs871514	4.12e-06	7	rs1395381	2.80e-06	2	rs4663333	2.58e-07
2	rs4663965	5.48e-07	2	rs4663333	5.97e-08	2	rs4663963	4.63e-06	3	rs9310867	4.57e-06	2	rs871514	2.87e-07
2	rs4399719	5.65e-07	2	rs4294999	6.26e-08	19	rs8108083	4.68e-06	12	rs7303705	4.87e-06	14	rs2353726	3.55e-07
2	rs3806597	6.46e-07	2	rs4663965	1.16e-07	6	rs199634	5.13e-06	4	rs3866838	5.14e-06	2	rs4294999	4.28e-07
2	rs7604115	6.46e-07	2	rs4663963	1.33e-07	19	rs2377572	5.64e-06	9	rs7847905	5.53e-06	2	rs4663965	5.99e-07

C Chapter IV (Creatinine Clearance association analyses)

Table 21 shows the pairwise correlation matrix of p-values from analysis of all subjects with baseline bilirubin as outcome

Table 21 Baseline bilirubin: Correlation matrix of p-values from combined group analysis

	CoCoBOT	**Additive**	**Dominant**	**Recessive**	**Categorical**
CoCoBOT	1				
Additive	0.813	1			
Dominant	0.72	0.736	1		
Recessive	0.204	0.319	0.066	1	
Categorical	0.568	0.647	0.65	0.646	1

Tables 22- 24 show the top 10 SNPs with the smallest p-values within each race/ethnicity group under each of the five analysis models with creatinine clearance as outcome

Table 22 6-month CrCl change: SNPs with the smallest p-values among African Americans

CoCoBOT			**Additive**			**Dominant**			**Recessive**			**Categorical**		
CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P
4	rs10518639	1.55e-06	13	rs4942321	1.00e-06	17	rs7224983	2.81e-07	18	rs9959038	1.50e-08	18	rs9959038	8.06e-08
23	rs6627929	6.21e-06	17	rs7224983	1.19e-06	16	rs10459845	3.88e-06	10	rs4934394	2.80e-08	13	rs7338174	1.08e-07
17	rs7224983	6.87e-06	12	rs2612060	3.05e-06	2	rs6435051	4.35e-06	13	rs7338174	7.35e-08	10	rs4934394	1.15e-07
15	rs12594783	6.98e-06	1	rs12070814	8.62e-06	7	rs9986688	4.94e-06	5	rs4957646	1.01e-07	18	rs629178	1.52e-07
1	rs12070814	7.35e-06	15	rs12594783	9.19e-06	8	rs6994092	6.21e-06	18	rs629178	1.06e-07	18	rs1443333	2.94e-07
9	rs2298181	8.18e-06	13	rs1327620	1.11e-05	3	rs6799661	8.51e-06	18	rs1443333	1.10e-07	7	rs6973776	3.52e-07
3	rs6792668	8.21e-06	16	rs10459845	1.40e-05	9	rs10812910	9.02e-06	14	rs2415485	1.31e-07	5	rs11960184	4.58e-07
2	rs17045635	8.36e-06	7	rs9986688	1.64e-05	9	rs10812913	1.02e-05	18	rs997105	1.63e-07	16	rs8063242	5.50e-07
10	rs2839658	8.85e-06	3	rs900370	1.66e-05	11	rs2291841	1.09e-05	6	rs3749930	2.52e-07	11	rs7950161	5.58e-07
7	rs9986688	9.45e-06	17	rs9908413	1.79e-05	13	rs1327620	1.40e-05	8	rs7003737	2.61e-07	11	rs7937105	5.58e-07

Table 23 6-month CrCl change: SNPs with the smallest p-values among Europeans

CoCoBOT			**Additive**			**Dominant**			**Recessive**			**Categorical**		
CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P
3	rs9820757	1.14e-06	5	rs1825192	7.27e-07	2	rs11682044	6.98e-07	4	rs16856919	5.02e-06	5	rs1825192	3.82e-06
6	rs977608	1.25e-06	3	rs7642361	2.73e-06	2	rs13388291	1.58e-06	4	rs4419508	5.02e-06	2	rs11682044	4.53e-06
10	rs10901541	1.77e-06	3	rs1604380	3.02e-06	2	rs13432998	1.65e-06	9	rs2905476	9.65e-06	3	rs7642361	5.75e-06
3	rs6808005	2.32e-06	8	rs4377920	3.52e-06	2	rs6545026	2.28e-06	8	rs7814638	1.11e-05	6	rs9388813	6.83e-06
3	rs2201439	2.45e-06	8	rs7003418	3.64e-06	6	rs977608	4.86e-06	8	rs6987706	1.15e-05	3	rs1604380	8.49e-06
8	rs4377920	2.64e-06	2	rs6545026	3.83e-06	2	rs11676168	4.90e-06	2	rs2723190	1.27e-05	2	rs13388291	8.62e-06
3	rs6550032	2.66e-06	8	rs10216910	4.26e-06	8	rs4732890	5.64e-06	2	rs2723168	1.28e-05	2	rs13432998	9.45e-06
8	rs7003418	3.52e-06	8	rs4732832	4.66e-06	8	rs4732657	5.64e-06	2	rs2708964	1.33e-05	2	rs6545026	9.89e-06
8	rs4732832	3.71e-06	5	rs2307116	7.25e-06	8	rs13280242	5.64e-06	2	rs2708965	1.34e-05	3	rs10804692	1.37e-05
8	rs10216910	3.72e-06	3	rs10804692	7.60e-06	3	rs9820757	6.26e-06	2	rs2708963	1.34e-05	5	rs10520873	1.40e-05

Table 24 6-month CrCl change: SNPs with the smallest p-values among Hispanics

CoCoBOT			Additive			Dominant			Recessive			Categorical		
CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P
2	rs10432654	6.38e-08	5	rs7712026	3.12e-06	5	rs7712026	1.65e-06	1	rs12406740	2.62e-09	19	rs4807371	2.01e-08
20	rs6054512	2.82e-06	10	rs942525	3.67e-06	13	rs9520594	1.75e-06	19	rs4474816	2.62e-09	19	rs4474816	2.07e-08
5	rs10071986	2.85e-06	20	rs6054512	4.77e-06	10	rs942525	3.37e-06	19	rs3746073	2.62e-09	19	rs3746073	2.09e-08
3	rs7641490	3.02e-06	13	rs9520594	5.83e-06	1	rs213045	4.62e-06	19	rs4807371	2.62e-09	1	rs12406740	2.10e-08
5	rs6895299	3.11e-06	12	rs2292249	7.32e-06	5	rs3776801	6.63e-06	19	rs7251272	2.82e-09	19	rs7251272	2.26e-08
10	rs942525	3.22e-06	1	rs1572507	1.03e-05	15	rs4984592	1.14e-05	3	rs264079	5.03e-08	9	rs7872379	1.57e-07
14	rs2807769	3.87e-06	8	rs11987198	1.09e-05	2	rs10432654	1.46e-05	22	rs2055183	6.75e-08	6	rs7741934	2.10e-07
19	rs2607416	4.25e-06	5	rs3776801	1.13e-05	1	rs7552569	1.68e-05	9	rs7872379	8.41e-08	3	rs264079	2.45e-07
7	rs10953236	5.19e-06	2	rs10432654	1.17e-05	20	rs6054512	2.08e-05	22	rs17001167	9.32e-08	22	rs2055183	4.23e-07
1	rs7552569	6.06e-06	10	rs7893939	1.29e-05	8	rs11987198	2.19e-05	6	rs13206561	1.49e-07	6	rs13206561	5.45e-07

Table 25 shows the top 20 SNPs with the smallest p-values in the analysis of all subjects under each of the five analysis models with baseline plasma bilirubin as outcome

Table 25 Baseline bilirubin: SNPs with the smallest p-values in combined group analysis

CoCoBOT			Additive			Dominant			Recessive			Categorical		
CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P	CHR	SNP	P
2	rs4148324	1.29e-15	2	rs6742078	4.68e-17	2	rs3755319	2.23e-13	2	rs4148325	1.84e-16	2	rs6742078	5.26e-18
2	rs6742078	1.41e-15	2	rs4148324	4.71e-17	2	rs4663333	9.43e-10	2	rs887829	2.35e-16	2	rs4148324	7.84e-18
2	rs887829	3.89e-15	2	rs887829	1.46e-16	2	rs4399719	9.70e-10	2	rs6742078	3.12e-16	2	rs887829	9.45e-18
2	rs4148325	7.57e-15	2	rs4148325	2.94e-16	2	rs4124874	1.57e-09	2	rs4148324	6.32e-16	2	rs4148325	1.17e-17
2	rs3755319	8.01e-14	2	rs10179091	2.47e-14	2	rs4663967	1.70e-09	2	rs10179091	4.55e-13	2	rs3755319	4.52e-14
2	rs10179091	1.42e-13	2	rs3755319	2.72e-14	2	rs4663965	1.92e-09	2	rs4148326	6.93e-13	2	rs10179091	4.99e-14
2	rs4148326	6.05e-13	2	rs4148326	2.18e-13	2	rs6431628	2.87e-09	2	rs3771341	4.46e-11	2	rs4148326	2.37e-13
2	rs3771341	1.10e-11	2	rs3771341	9.72e-13	2	rs4148324	2.93e-09	2	rs929596	1.27e-10	2	rs3771341	6.31e-13
2	rs929596	5.80e-11	2	rs871514	1.71e-12	2	rs6742078	3.61e-09	2	rs17862875	3.15e-10	2	rs929596	1.87e-12
2	rs871514	6.93e-11	2	rs929596	3.19e-12	2	rs2008595	3.89e-09	2	rs2221198	2.45e-09	2	rs871514	1.54e-11
2	rs4294999	9.43e-11	2	rs4663965	3.22e-12	2	rs7572563	3.95e-09	2	rs4663969	2.81e-09	2	rs17862875	1.73e-11
2	rs4663965	1.09e-10	2	rs4294999	3.40e-12	2	rs3806597	4.39e-09	2	rs7556676	3.29e-09	2	rs4663965	2.69e-11
2	rs4663333	1.10e-10	2	rs4663333	4.80e-12	2	rs4294999	5.49e-09	2	rs7604115	6.26e-09	2	rs4294999	3.05e-11
2	rs4399719	1.79e-10	2	rs4663963	7.28e-12	2	rs17862866	6.47e-09	2	rs871514	6.60e-09	2	rs4663333	3.40e-11
2	rs2221198	1.83e-10	2	rs4663967	8.71e-12	2	rs4663963	7.21e-09	2	rs4294999	4.84e-08	2	rs4663967	6.22e-11
2	rs7556676	2.37e-10	2	rs3806597	1.15e-11	2	rs871514	8.32e-09	2	rs4663963	8.88e-08	2	rs4663963	6.37e-11
2	rs4663967	2.38e-10	2	rs4399719	1.73e-11	2	rs887829	1.63e-08	2	rs4663965	9.11e-08	2	rs3806597	9.30e-11
2	rs4663963	2.42e-10	2	rs6431628	3.23e-11	2	rs4148325	3.57e-08	5	rs35139949	1.05e-07	2	rs4399719	9.46e-11
2	rs4663969	2.64e-10	2	rs4124874	3.39e-11	2	rs3771341	1.00e-07	5	rs35981677	1.34e-07	2	rs4124874	1.81e-10
2	rs3806597	2.65e-10	2	rs2008595	3.79e-11	2	rs10179091	1.13e-07	2	rs3806597	1.66e-07	2	rs6431628	2.04e-10

REFERENCES

[1] Agresti A. Considerations in measuring partial association for ordinal categorical data. Journal of the American Statistical Association: 1977;72(357):3745.

[2] Agresti A. Categorical Data Analysis (2nd ed.), Hoboken, New Jersey: John Wiley: 2002.

[3] Ananth CV and Kleinbaum DG. Regression models for ordinal responses: A review of methods and applications. International Journal of Epidemiology: 1997;26(6):1323-1333.

[4] Barlow RE, Bartholomew DJ, Bremner JM, Brunk HD. Statistical Inference Under Order Restrictions. London: John Wiley: 1972

[5] Davis JA. A partial coefficient for Goodman and Kruskals gamma. Journal of the American Statistical Association: 1967;62(317):189193.

[6] Goodman LA. Simple models for the analysis of association in cross-classifications having ordered categories. Journal of the American Statistical Association: 1979; 74(367):537-552.

[7] Hawkes RK. The multivariate analysis of ordinal measures. American Journal of Sociology: 1971;76(5):908926.

[8] Kendall, MG. (1948), Rank Correlation Methods, Hafner: New York: 1948.

[9] Li C, Shepherd BE. Test of association between two ordinal variables while adjusting for covariates. Journal of the American Statistical Association: 2010;105(490):612620.

[10] Mantel N. Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. Journal of the American Statistical Association: 1963;58(303):690700.

[11] Ramsay JO. Monotone regression splines in action. Statistical Science: 1988;3(4):425-461.

[12] Stefanski LA, Boos DD. The calculus of M-estimation. The American Statistician: 2002;56(1): 2938.

[13] Torgerson WS. A non-parametric test of correlation using rank orders within subgroups. Psychometrika: 1956;21(2):145152.

[14] Walter SD, Feinstein AR, Well CK. Coding ordinal independent variables in multiple regression analyses. American Journal of Epidemiology: 1987;125(2):319-323.

[15] The Panel on Clinical Practices for Treatment of HIV Infection. Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents. Available at: http://www.aidsinfo.nih.gov/contentfiles/lvguidelines/adultandadolescentgl.pdf. Accessed Updated February 21, 2014.

[16] Kearney BP, Flaherty JF, Shah J. Tenofovir disoproxil fumarate: clinical pharmacology and pharmacokinetics. ClinPharmacokinet 2004;43(9):595-612.

[17] Gallant JE, Staszewski S, Pozniak AL, et al. Efficacy and safety of tenofovir DF vs stavudine in combination therapy in antiretroviral-naive patients: a 3-year randomized trial. JAMA : the journal of the American Medical Association 2004 Jul 14;292(2):191-201.

[18] Gallant JE, DeJesus E, Arribas JR, et al. Tenofovir DF, emtricitabine, and efavirenz vs. zidovudine, lamivudine, and efavirenz for HIV. NEnglJ Med 2006;354(3):251-60.

[19] Cassetti I, Madruga JV, Suleiman JM, et al. The safety and efficacy of tenofovir DF in combination with lamivudine and efavirenz through 6 years in antiretroviral-naive HIV-1-infected patients. HIV Clinical Trials 2007 May-Jun;8(3):164-72.

[20] Molina JM, Andrade-Villanueva J, Echevarria J, et al. Once-daily atazanavir/ritonavir versus twice-daily lopinavir/ritonavir, each in combination with tenofovir and emtricitabine, for management of antiretroviral-naive HIV-1-infected patients: 48 week efficacy and safety results of the CASTLE study. Lancet 2008 Aug 23;372(9639):646-55.

[21] Ortiz R, Dejesus E, Khanlou H, et al. Efficacy and safety of once-daily darunavir/ritonavir versus lopinavir/ritonavir in treatment-naive HIV-1-infected patients at week 48. AIDS 2008 Jul 31;22(12):1389-97.

[22] Sax PE, Tierney C, Collier AC, et al. Abacavir-lamivudine versus tenofovir-emtricitabine for initial HIV-1 therapy. N Engl J Med 2009 Dec 3;361(23):2230-40.

[23] Lennox JL, DeJesus E, Lazzarin A, et al. Safety and efficacy of raltegravir-based versus efavirenz-based combination therapy in treatment-naive patients with HIV-1 infection: a multicentre, double-blind randomised controlled trial. Lancet 2009 Sep 5;374(9692):796-806.

[24] Post FA, Moyle GJ, Stellbrink HJ, et al. Randomized comparison of renal effects, efficacy, and safety with once-daily abacavir/lamivudine versus tenofovir/emtricitabine, administered with efavirenz, in antiretroviral-naive, HIV-1-infected adults: 48-week results from the ASSERT study. Journal of Acquired Immune Deficiency Syndromes 2010 Sep;55(1):49-57.

[25] Daar ES, Tierney C, Fischl MA, et al. Atazanavir plus ritonavir or efavirenz as part of a 3-drug regimen for initial treatment of HIV-1. Ann Intern Med 2011 Apr 5;154(7):445-56.

[26] Rockstroh JK, Lennox JL, Dejesus E, et al. Long-term treatment with raltegravir or efavirenz combined with tenofovir/emtricitabine for treatment-naive human immunodeficiency virus-1-infected patients: 156-week results from STARTMRK. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 2011 Oct;53(8):807-16.

[27] Cohen CJ, Molina JM, Cahn P, et al. Efficacy and safety of rilpivirine (TMC278) versus efavirenz at 48 weeks in treatment-naive HIV-1-infected patients: pooled results from the phase 3 double-blind randomized ECHO and THRIVE Trials. Journal of Acquired Immune Deficiency Syndromes 2012 May 1;60(1):33-42.

[28] Gerard L, Chazallon C, Taburet AM, Girard PM, Aboulker JP, Piketty C. Renal function in antiretroviral-experienced patients treated with tenofovir disoproxil fumarate associated with atazanavir/ritonavir. Antivir Ther 2007;12(1):31-9.

[29] Goicoechea M, Liu S, Best B, et al. Greater tenofovir-associated renal function decline with protease inhibitor-based versus nonnucleoside reverse-transcriptase inhibitor-based therapy. J Infect Dis 2008 Jan 1;197(1):102-8. 16.

[30] Horberg M, Tang B, Towner W, et al. Impact of Tenofovir on Renal Function in HIV-Infected Antiretroviral Naive Patients. J Acquir Immune Defic Syndr 2009 Oct 15.

[31] Barditch-Crovo P, Deeks SG, Collier A, et al. Phase i/ii trial of the pharmacokinetics, safety, and antiretroviral activity of tenofovir disoproxil fumarate in human immunodeficiency virus-infected adults. Antimicrobial agents and chemotherapy 2001 Oct;45(10):2733-9.

[32] Ray AS, Cihlar T, Robinson KL, et al. Mechanism of active renal tubular efflux of tenofovir. Antimicrobial agents and chemotherapy 2006 Oct;50(10):3297-304.

[33] Izzedine H, Hulot JS, Villard E, et al. Association between ABCC2 gene haplotypes and tenofovir-induced proximal tubulopathy. The Journal of infectious diseases 2006 Dec 1;194(11):1481-91.

[34] Nishijima T, Komatsu H, Higasa K, et al. Single nucleotide polymorphisms in ABCC2 associate with tenofovir-induced kidney tubular dysfunction in Japanese patients with HIV-1 infection: a pharmacogenetic study. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 2012 Dec;55(11):1558-67.

[35] Pushpakom SP, Liptrott NJ, Rodriguez-Novoa S, et al. Genetic variants of ABCC10, a novel tenofovir transporter, are associated with kidney tubular dysfunction. The Journal of infectious diseases 2011 Jul 1;204(1):145-53.

[36] Kiser JJ, Aquilante CL, Anderson PL, King TM, Carten ML, Fletcher CV. Clinical and genetic determinants of intracellular tenofovir diphosphate concentrations in HIV-infected patients. Journal of Acquired Immune Deficiency Syndromes 2008 Mar 1;47(3):298-303.

[37] Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. Nephron 1976;16(1):31-41.

[38] Haas DW, Wilkinson GR, Kuritzkes DR, et al. A multi-investigator/institutional DNA bank for AIDS-related human genetic studies: AACTG Protocol A5128. HIV Clin Trials 2003 Sep-Oct;4(5):287-300.

[39] Pereyra F, Jia X, McLaren PJ, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 2010 Dec 10;330(6010):1551-7.

[40] Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007 Sep;81(3):559-75.

[41] Olagunju A, Owen A, Cressey TR. Potential effect of pharmacogenetics on maternal, fetal and infant antiretroviral drug exposure during pregnancy and breastfeeding. Pharmacogenomics 2012 Oct;13(13):1501-22.

[42] Michaud V, Bar-Magen T, Turgeon J, Flockhart D, Desta Z, Wainberg MA. The dual role of pharmacogenetics in HIV treatment: mutations and polymorphisms regulating antiretroviral drug resistance and disposition. Pharmacol Rev 2012 Jul;64(3):803-33.

[43] Rodriguez-Novoa S, Labarga P, Soriano V, et al. Predictors of kidney tubular dysfunction in HIV-infected patients treated with tenofovir: a pharmacogenetic study. Clin Infect Dis 2009 Jun 1;48(11):e108-16.

[44] Kiser JJ, Carten ML, Aquilante CL, et al. The effect of lopinavir/ritonavir on the renal clearance of tenofovir in HIV-infected patients. Clinical pharmacology and therapeutics 2008 Feb;83(2):265-72.

[45] Rodriguez-Novoa S, Labarga P, Soriano V. Pharmacogenetics of tenofovir treatment. Pharmacogenomics 2009 Oct;10(10):1675-85.

[46] Bleasby K, Hall LA, Perry JL, Mohrenweiser HW, Pritchard JB. Functional consequences of single nucleotide polymorphisms in the human organic anion transporter hOAT1 (SLC22A6). The Journal of pharmacology and experimental therapeutics 2005 Aug;314(2):923-31.

[47] Yee SW, Chen L, Giacomini KM. Pharmacogenomics of membrane transporters: past, present and future. Pharmacogenomics Apr;11(4):475-9.

[48] Hirt D, Urien S, Ekouevi DK, et al. Population pharmacokinetics of tenofovir in HIV-1-infected pregnant women and their neonates (ANRS 12109). Clin Pharmacol Ther 2009 Feb;85(2):182-9.

[49] Hindorff LA, MacArthur J. (European Bioinformatics Institute), Morales J (European Bioinformatics Institute), Junkins HA, Hall PN, Klemm AK, and Manolio TA. A Catalog of Published Genome-Wide Association Studies. Available at: http://www.genome.gov/gwastudies. Accessed September 19, 2013.

[50] Venuto CS, Ma Q, Daar ES, Sax PE, Fischl MA, Collier AC, Mollan K, Smith K, Tierney C, Morse GD, The ACTG 5202 Protocol Team. Tenofovir Plasma Pharmacokinetics in the AIDS Clinical Trials Group Study A5202. Poster Exhibition, 19th International AIDS Conference, Washington DC, USA, 2012; Poster TUPE058

[51] King JR, Yogev R, Jean-Philippe P, Graham B, Wiznia A, Britto P, Carey V, Hazra R, Acosta EP, P1058 Protocol Team. Steady-state pharmacokinetics of tenofovir-based regimens in HIV-infected pediatric patients. Antimicrob Agents Chemother 2011; 55(9): 4290-4294.

[52] Patrias K, author; Wendling D, editor. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers [Internet]. 2nd edition. Bethesda (MD): National Library of Medicine (US); 2007-. Chapter 24, Databases/Retrieval Systems on the Internet. 2007 Oct 10 [Updated 2011 Sep 15]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK7273/

[53] Mengjin Z, Shuhong Z, Candidate Gene Identification Approach: Progress and Challenges. International Journal of Biology 2007 October;3(7):420-427.

[54] Johnson A.D, Handsaker RE, Pulit S, Nizzari MM, O'Donnell C J, de Bakker PIW. SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap Bioinformatics, 2008 24(24):2938-2939

[55] McDonagh EM, Whirl-Carrillo M, Garten Y, Altman RB, Klein TE, From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource. Biomarkers in Medicine 2011; 5(6):795-806

[56] Pruim RJ*, Welch RP*, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. (2010) LocusZoom: Regional visualization of genome-wide association scan results. Bioinformatics 2010 September 15; 26(18): 2336.2337.

[57] Breiderhoff T, Himmerkus N, Stuiver M, Mutig K, Will C, Meij IC, Bachmann S, Bleich M, Willnow TE, Muller D. Deletion of claudin-10 (Cldn10) in the thick ascending limb impairs paracellular sodium permeability and leads to hypermagnesemia and nephrocalcinosis. National Academy of Sciences 2012; 109(35):14241-14246.

[58] Reznichenko A, Sinkeler SJ, Snieder H, van den Born J, de Borst MH, Damman J, van Dijk MC, van Goor H, Hepkema BG, Hillebrands JL, Leuvenink HG, Niesing J, Bakker SJ, Seelen M, Navis G. SLC22A2 is associated with tubular creatinine secretion and bias of estimated GFR in renal transplantation. Physiol Genomics 2013; 45(6):201-209

[59] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 2006; 38: 904-909

[60] Johnson DH, Venuto C, Ritchie MD, Morse GD, Daar ES, McLaren PJ, Haas DW (2014). Genomewide association study of atazanavir pharmacokinetics and hyperbilirubinemia in AIDS Clinical Trials Group protocol A5202. Pharmacogenet Genomics 24: 195-203.

[61] Baheti G, Kiser JJ, Havens PL, Fletcher CV. Plasma and intracellular population pharmacokinetic analysis of tenofovir in HIV-1-infected patients. Antimicrob Agents Chemother 2011; 55(11): 5294-5299

[62] Holzinger ER, Grady B, Ritchie MD, Ribaudo HJ, Acosta EP, Morse GD, Gulick RM, Robbins GK, Clifford DB, Daar ES, McLaren P, Haas DW (2012) Genome-Wide Association Study of Plasma Efavirenz Pharmacokinetics in AIDS Clinical Trials Group Protocols. Pharmacogenet Genom 22: 858-67.

[63] Nelson MR, Bacanu SA, Mosteller M, Li L, Bowman CE, Roses AD, Lai EH, Ehm MG (2009). Genome-wide approaches to identify pharmacogenetic contributions to adverse drug reactions. The pharmacogenomics journal 9: 23-33

[64] Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, Gut I, Lathrop M, Collins R (2008). SLCO1B1 variants and statin-induced myopathy–a genomewide study. NEnglJ Med 359: 789-799