

USE OF BIOLOGICAL KNOWLEDGE TO INFORM THE ANALYSIS OF GENE-GENE INTERACTIONS INVOLVED IN MODULATING VIROLOGIC FAILURE WITH EFAVIRENZ-CONTAINING TREATMENT REGIMENS IN ART-NAÏVE ACTG CLINICAL TRIALS PARTICIPANTS

BENJAMIN J. GRADY¹, ERIC S. TORSTENSON¹, PAUL J. MCLAREN², PAUL I.W. DE BAKKER², DAVID W. HAAS³, GREGORY K. ROBBINS⁴, ROY M. GULICK⁵, RICHARD HAUBRICH⁶, HEATHER RIBAUDO⁷ AND MARYLYN D. RITCHIE^{1*}

¹*Center for Human Genetics Research, Vanderbilt University, Nashville, TN 37232, USA*

²*Broad Institute, Harvard University, Cambridge, Massachusetts 02138, USA*

³*Departments of Medicine, Microbiology & Immunology, Vanderbilt University, Nashville, TN 37232, USA*

⁴*Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA*

⁵*Department of Medicine, Cornell University, New York, NY 10065, USA*

⁶*Department of Medicine, University of California, San Diego, San Diego, CA 92103, USA*

⁷*Department of Biostatistics, Harvard University, Boston, MA 02115, USA*

**Corresponding Author Email: ritchie@chgr.mc.vanderbilt.edu*

Personalized medicine is a high priority for the future of health care. The idea of tailoring an individual's wellness plan to their unique genetic code is one which we hope to realize through the use of pharmacogenomics. There have been examples of tremendous success in pharmacogenomic associations however there are many such examples in which only a small proportion of trait variance has been explained by the genetic variation. Although the increased use of GWAS could help explain more of this variation, it is likely that a significant proportion of the genetic architecture of these pharmacogenomic traits are due to complex genetic effects such as epistasis, also known as gene-gene interactions, as well as gene-drug interactions. In this study, we utilize the Biofilter software package to look for candidate epistasis contributing to risk for virologic failure with efavirenz-containing antiretroviral therapy (ART) regimens in treatment-naïve participants of AIDS Clinical Trials Group (ACTG) randomized clinical trials. A total of 904 individuals from three ACTG trials with data on efavirenz treatment are analyzed after race-stratification into white, black, and Hispanic ethnic groups. Biofilter was run considering 245 candidate ADME (absorption, distribution, metabolism, and excretion) genes and using database knowledge of gene and protein interaction networks to produce approximately 2 million SNP-SNP interaction models within each ethnic group. These models were evaluated within the PLATO software package using pair wise logistic regression models. Although no interaction model remained significant after correction for multiple comparisons, an interaction between SNPs in the *TAPI* and *ABCC9* genes was one of the top models before correction. The *TAPI* protein is responsible for intracellular transport of antigen to MHC class I molecules, while *ABCC9* codes for a transporter which is part of the subfamily of ABC transporters associated with multi-drug resistance. This study demonstrates the utility of the Biofilter method to prioritize the search for gene-gene interactions in large-scale genomic datasets, although replication in a larger cohort is required to confirm the validity of this particular *TAPI-ABCC9* finding.

1. Introduction

1.1. *The HIV pandemic*

Human Immunodeficiency Virus (HIV) Type 1 infection has been in a state of pandemic for several years. The 2008 UNAID report estimates that 33 million people are currently infected, with approximately 3 million new infections during the year of 2008 [1]. Within regions in Sub-Saharan Africa, the prevalence of HIV-1 infection rises as high as 25-30% [1]. Because there is no cure for HIV-1 infection, one of the best tools available to combat the epidemic currently is antiretroviral therapy (ART). ART helps with treating those individuals already infected and helps to reduce the chance of spreading the disease[2]. ART consists of a regimen of two or three antiretroviral drugs and is successful in drastically increasing the lifespan of HIV-1 infected individuals and improving their quality of life[3]. By reducing the amount of virus circulating freely in the blood of an infected person, ART also greatly decreases the probability of transmitting the virus through sexual contact[4], and child birth [5]. Despite the benefits of ART for its use in fighting HIV, there are unfortunately several issues that accompany the use of the therapy. Arguably the most significant issue among these is the prevalence of adverse drug reactions (ADR) and the failure of the drug to suppress viral load. Adverse reactions to antiretroviral drugs range from skin rash and nausea to neurologic impairment and fatal hypersensitivity, as is sometimes seen in response to the drug abacavir [6]. ADRs contribute to ineffectiveness of ART by reducing adherence to drug regimens and requiring temporary discontinuation of treatment [7]. The failure of a drug to suppress viral load in a patient is known as virologic failure [8]. Virologic failure refers either to initial inefficacious response to the drug and a failure to ever reach a controlled viral load or to the phenomenon whereby viral load rebounds subsequent to reaching a controlled level.

1.2. *Pharmacogenomics and HIV treatment*

The way in which people respond to drug treatment has been shown, in many cases, to be influenced by their genetics. The field of pharmacogenomics attempts to discover the exact genetic variants which predict success, failure or ADR in response to treatment. There have been successes in identifying genetic polymorphisms which explain large proportions of variance in drug response. Approximately 20-30% of the variance in initial dosing of the anti-coagulant warfarin, for example, can be explained by variation in the gene *VKORC1*[9], which codes for vitamin K epoxide reductase complex subunit 1. Vitamin K epoxide reductase creates the enzymatically active form of vitamin K [10] which is in turn extremely important in modulating the function of proteins involved in blood clotting. For this reason, it makes biological sense that a polymorphism which affects the expression of *VKORC1* would also affect how much warfarin is necessary to prevent over-clotting. Arguably the most significant pharmacogenomic discovery has been made in the field of HIV ART. Hypersensitivity reaction (HSR) in response to the nucleoside reverse-transcriptase inhibitor (NRTI) abacavir, a commonly-used drug in ART regimens, has been shown to be strongly tied to *HLA*

genotype. *HLA-B*5701* genotype has a 100% negative predictive value (NPV) for predicting HSR from abacavir [11, 12]. As a result of this relationship, *HLA-B*5701* has become one of the first genetic tests approved by the FDA for use in determining risk prior to prescription of a drug. Although the abacavir story represents the pinnacle of pharmacogenomic discovery and there may not be another single genetic polymorphism with 100% NPV for ART in the future, there are still many possibilities for utilizing genetic prediction models in determination of the optimal ART drug regimen to prescribe in order to control HIV. It might be that a combination of genetic variants in concert would best predict antiretroviral drug response.

1.3. Genetic interactions

Decades of research into the pharmacokinetics of drug metabolism have shown that the enzymes which process and transport pharmaceuticals function as part of highly-interconnected networks [13]. For example, studies have shown that many drugs, including phenytoin [14] and irinotecan [15], can be metabolized, activated, or deactivated by more than one enzyme [16]. It is as a result of this complementation that it is reasonable to expect the necessity of multiple genetic polymorphisms to experience a large change in the resulting phenotype. The phenomenon of gene-gene interaction, or epistasis as it is often referred to in the field of genetic epidemiology, has been a subject of much discussion over the past decade [17-19]. Although the term epistasis was coined separately by Bateson [20] and Fisher [21] in the early 20th century to refer to the effect of one gene “masking” another’s effect or a non-additive effect of multiple elements observed simultaneously, respectively, the necessary technology to explore its presence has only recently been developed. The HapMap project, the sequencing of the human genome, and the steady increase in computational power have been the driving factors in the ability to analyze genetic data for gene-gene interaction effects. Despite the rising computational power available, genotyping technology has far out-paced the ability to exhaustively analyze multi-locus genetic effects for genome-wide association study (GWAS) data. To search exhaustively for epistasis between two single nucleotide polymorphisms (SNPs) in a current GWAS containing 1 million SNPs would require 5×10^{11} tests. Although it is still possible to perform this pair wise exploration by utilizing parallel computation, it is clear that with the advent of whole-exome and whole-genome sequencing technology as a primary source for genetic information in association studies in the near future, an alternative to exhaustive searches must be found. One such solution is that of biasing the search using prior knowledge to search for combinations of genes that are likely to interact biologically. The Biofilter tool [22] was developed to use databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the Protein Family database (PFAM) in order to build SNP-SNP models based on known interactions between genes and proteins in curated pathways and networks. Especially in a field such as pharmacogenomics, in which the knowledge of the drug metabolism networks is extensive, enriching the search for epistasis with knowledge from known biological interactions could prove valuable. Not only does this alleviate the issues of computational complexity, but it also substantially reduces the number of tests and associated

multiple-comparison issues. As opposed to considering 10's or 100's of billions of two-way interaction models, one would search a more reasonable subset of a few million models with a solid biological basis. The time necessary to perform the subsequent statistical analysis declines from days to hours, using a single processor. One concern of pursuing only a biologically-informed subset of interaction models is the possible loss of novel significant interactions during filtering. Due to long-standing knowledge in pharmacology, the potential reduction in noise outweighs the concern.

2. Methods

2.1. Study population

DNA samples in the current study come from individuals who were randomized to receive efavirenz (in multidrug ART regimens) in the AIDS Clinical Trials Group (ACTG) randomized clinical trials (RCT) ACTG 384, A5095 and A5142 and were collected under protocol A5128 [23]; study designs are described in depth elsewhere [24-30]. ACTG 384 [29, 30] and A5095 [24-26] were double-blind, multicenter RCTs designed to test the effectiveness of differing ART drug regimens. Of the 980 individuals enrolled into ACTG 384, 526 were consented for DNA extraction and 347 of those with DNA available were on efavirenz-containing regimens. A5095 enrolled 1147 subjects for comparison of protease inhibitor-sparing regimens. Of the enrollment in A5095, a total of 600 individuals were available who both consented for DNA and had ART containing efavirenz. The final study used in this multi-study analysis was A5142 [27, 28]. This ACTG study was a Phase III comparison of 3 ART regimens. Of the 757 participants of A5142, 411 were randomized to receive ART containing efavirenz and provided DNA samples.

Table 1. Sample size broken down by study and race ethnicity grouping.

			Study			
			ACTG 384	A5095	A5142	Total
Pre-QC	Total	Cases	45	124	97	266
		Controls	302	476	314	1092
Post-QC	Total	Cases	38	100	59	197
		Controls	228	319	160	707
	White	Cases	16	34	24	74
		Controls	116	163	78	357
	Black	Cases	18	47	30	95
		Controls	69	97	49	215
	Hispanic	Cases	4	19	5	28
		Controls	43	59	33	135

The total number of individuals available with DNA samples and GWAS genotyping across these three studies was 1358. Self-described ethnicity of the combined study population reveals the study population to consist of 45% white (N = 606), 34% non-Hispanic black (N = 459), 19% Hispanic (N = 265) and 2% other (N = 28). After quality control (QC) and exclusions were applied, 904 participants remained available for analysis. Of this 904, 48% (N = 431) were non-Hispanic white, 34% were non-Hispanic black (N = 310), and 18% (N = 163) were Hispanic. The endpoint used in this study was virologic failure as defined by a spike in viral load above 200 copies/mL after achieving viral load less than 200 copies/mL on ART. Individuals who experienced virologic failure on efavirenz are categorized as cases while those who did not are categorized as controls (**Table 1**).

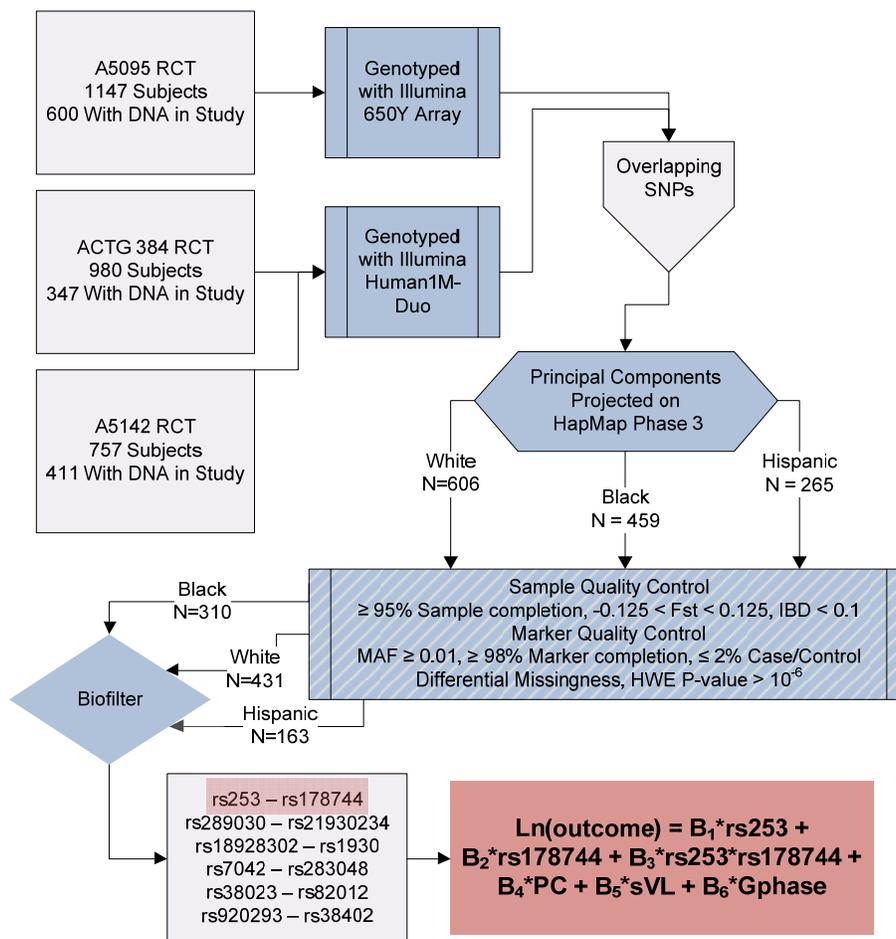


Figure 1. An outline of the analysis plan used in this study. A more detailed description of the Biofilter step is available in Figure 2.

2.2. Genotyping and quality control exclusions

Individuals from the ACTG 384 study were genotyped on the Illumina 650Y array while those from A5142 and A5095 were genotyped with the Illumina Human1M-Duo platform (Illumina, Inc. San Diego, CA). In combining data from these two platforms, only the SNPs which overlapped were used for this analysis. Principal components analysis was performed referencing the HapMap phase 3 sample data to map each individual back to one of three major ethnic groups: white, black, and Hispanic. There was greater than 95% concordance between self-reported ethnicity and that found through principal components analysis. Within each race stratum, quality control was performed to filter out samples and SNPs of low quality (Figure 1). Samples with low genotyping rate (<95%), high or low heterozygosity (inbreeding coefficient > 0.125 or < -0.125) and related individuals (IBD estimate > 0.1) were removed. SNPs with missingness $> 2\%$, large deviations from Hardy-Weinberg equilibrium ($p < 10^{-6}$) and those with differential missingness between cases and controls $> 2\%$ were removed from analysis.

2.3. Biofilter

The Biofilter[22] was developed to provide prior biological knowledge to influence the search for gene-gene interactions in large-scale data. Given a set of variants, Biofilter first maps the SNPs back to genes based on gene definitions in Ensembl and then builds models using disease-dependent (i.e. those biological associations previously known with respect to the trait under investigation) or disease-independent relationships (i.e. known biological interactions with no particular association to the trait under consideration). A unique option for the Biofilter is to provide a personally curated list of genes based on expert knowledge of the phenotype under study as a starting point, using that list to search both disease-dependent and disease-independent data sources to map all other genes that are related to the genes in the curated list. Based on user options, the Biofilter can query the set of databases, which currently includes KEGG, PFAM, Reactome, DIP, PFAM, GO, and NetPath, to establish groups of interacting genes. Once these groups are established, SNP-SNP interaction models are created by exhaustively pairing two SNPs from two genes in the group. Biofilter allows flexibility in choosing how restrictive the creation of interaction models will be. For example, when inputting a list of self-curated genes, the user has the option to ensure that at least one of the SNPs comes from a gene in the list. Alternatively, restrictions can be relaxed to allow models with SNPs from other genes in the same group or even pathway as those genes in the list. As shown in Figure 2, we provided a list of 245 absorption, distribution, metabolism and elimination (ADME) genes which were curated by the authors and allowed for the inclusion of SNPs which were within 10kb of the gene boundaries. Interactions were restricted to allow only those models for which at least one of the SNPs in the model belonged to a gene in the list – although the search was conducted in the disease-independent databases. Two-SNP interaction models were generated separately for non-Hispanic white (henceforth referred to as white), non-Hispanic black (henceforth referred to as black), and

Hispanic ethnic groups, while those participants self-describing as other races were excluded. We used all six databases currently integrated in the Biofilter to generate our SNP-SNP models.

2.4. Statistical analysis

All statistical analyses were performed using the Platform for the Analysis, Translation and Organization of large scale data (PLATO) software package (<http://chgr.mc.vanderbilt.edu/ritchie/submit/subscriptions>) [31]. PLATO is a scaffold which allows for recoding, quality control, and analysis of data as part of a pipeline. The Biofilter models were used as input for PLATO. The statistical analyses performed used logistic regression to assess the risk of each pair of interaction models. Logistic regression models included terms for each SNP separately and a term for multiplicative interaction. In addition, variables deemed important with respect to the outcome of virologic failure were included as covariates in the model. Principal components vectors were utilized to adjust for population substructure within each racial group, as might exist between northern and southern European white individuals or African Americans. Indicator variables for genotyping phase and baseline viral load (\geq or $<$ 100,000 copies/mL) were also incorporated. Regression analysis was performed separately within each ethnic group as defined by principal components analysis.

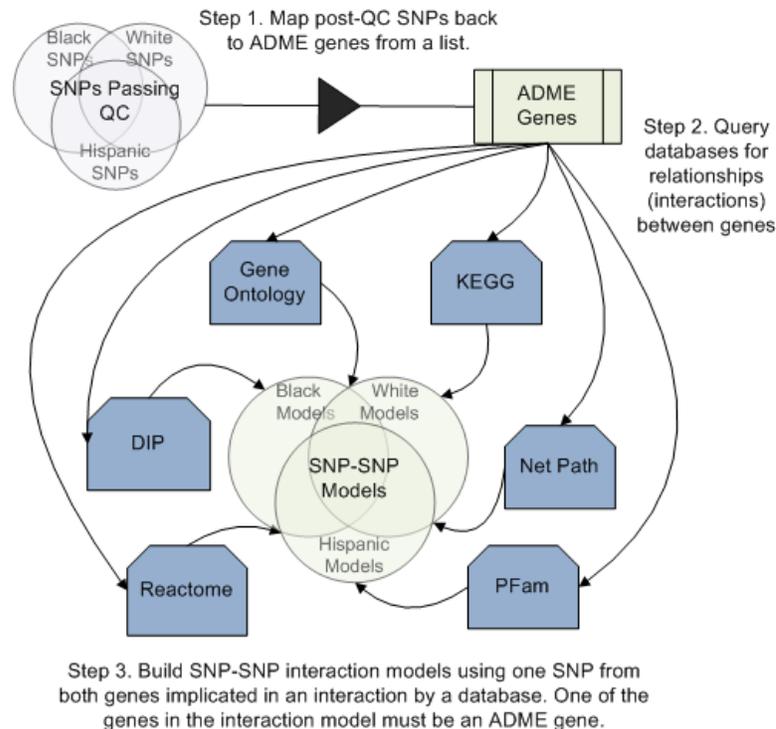


Figure 2. A schematic of the procedure involved in a Biofilter run.

3. Results

Genome-wide genotyping of 1358 AIDS Clinical Trials Group (ACTG) participants with exposure to the NNRTI efavirenz was conducted to elucidate the genetic basis of virologic failure. Race-stratification was performed using principal components analysis based on HapMap phase 3 samples. After quality control processes, 904 individuals remained. The Biofilter software tool was used to take a list of 245 ADME genes and build putative gene-gene interactions based on biological knowledge provided by KEGG, DIP, Pfam, Net Path, Reactome, and Gene Ontology. The SNPs from each ethnic group were then mapped back to these ADME genes and SNP-SNP models were created by taking one SNP from each gene in a proposed gene-gene interaction. Running Biofilter resulted in 2,144,157 models to evaluate in whites, 2,471,201 models in blacks, and 2,099,614 models in Hispanics. These models were derived from a total of 33067, 35764 and 32698 SNPs for white, black and Hispanic groups respectively. If all two-way interactions between these SNPs were exhaustively tested, it would result in the evaluation of 546 million models for the white group and 638 million and 534 million models for the black and Hispanic groups respectively. The differences in model number between ethnic groups are due to race-stratified quality control. SNP-SNP models from Biofilter were passed to PLATO[31] to perform logistic regression analysis. Due to the highly correlated nature of many of the interaction models, a Bonferroni correction would too conservative for correcting for multiple testing. Instead, a false-discovery rate (FDR) correction was applied using the qvalue package available in R. No interaction models were found to be significant at an FDR level of 0.10, although the most significant interactions were significant at an FDR level of 0.45. The interaction models with lowest p-values are shown in Table 2.

Table 2. Most significant interaction models resulting from gene-gene interaction analysis.

SNP1	SNP2	Model P-value	Interaction P-value	SNP1 Odds Ratio	SNP2 Odds Ratio	Interaction Odds Ratio
rs2318785 (<i>NME2</i>)	rs2157597 (<i>NME7</i>)	1.69E-06	5.98E-07	0.257	0.253	4.732
rs2318785 (<i>NME2</i>)	rs12118611 (<i>NME7</i>)	1.69E-06	5.98E-07	0.253	0.257	4.732
rs2318785 (<i>NME2</i>)	rs12121994 (<i>NME7</i>)	1.68E-06	6.15E-07	0.247	0.263	4.769
rs2318785 (<i>NME2</i>)	rs17349439 (<i>NME7</i>)	3.30E-06	1.20E-06	0.265	0.277	4.487
rs2318785 (<i>NME2</i>)	rs6703463 (<i>NME7</i>)	1.48E-05	3.93E-06	0.258	0.289	3.884
rs2318785 (<i>NME2</i>)	rs12744184 (<i>NME7</i>)	7.46E-06	6.46E-06	0.253	0.306	3.821
rs735883 (<i>TAP1</i>)	rs1283807 (<i>ABCC9</i>)	4.38E-06	9.05E-06	2.245	3.236	0.154
rs735883 (<i>TAP1</i>)	rs1352909 (<i>ABCC9</i>)	4.61E-06	9.76E-06	2.231	3.205	0.155
rs735883 (<i>TAP1</i>)	rs4148665 (<i>ABCC9</i>)	4.83E-06	9.89E-06	2.230	2.850	0.172
rs735883 (<i>TAP1</i>)	rs1283798 (<i>ABCC9</i>)	1.04E-05	1.00E-05	2.395	2.482	0.240

4. Discussion

As genotyping technologies progress and we move into the era of whole-genome sequencing, the need to improve analysis schemes is ever-present. This is especially true when gene-gene, gene-environment, and gene-drug interactions are concerned. Allowing our biological knowledge of gene and protein network dynamics to guide the search for the genetic basis of disease is a promising solution to this dilemma. While our current state of biological knowledge is limited, and that knowledge-base will continue to grow and develop over time, if we develop techniques that use the information we have, while still exploring novel interactions, we have a greater chance for success. By narrowing the dimensions of the search space, the computational complexity of the problem becomes much more amenable to current analytical techniques. In addition, interpretation of results is more straightforward. We utilized a list of 245 genes involved in absorption, distribution, metabolism and elimination of drugs and their metabolites to focus the search for gene-gene interactions associated with virologic failure during HIV treatment with efavirenz. Although there were no gene-gene interactions which remained significant after correction for multiple testing, this could be related to the small sample size present in this study. Due to race-stratification, the largest group in the analysis had 74 cases and 357 controls. But the development of this analytic pipeline and software tools will be immensely useful for future analyses.

The interactions which appeared most significant in the results of the logistic regression analysis occur between a SNP - rs2318785 - in the *NME2* gene and multiple SNPs in the *NME7* gene. Both *NME2* and *NME7* are part of the NDK family, coding for nucleoside diphosphate kinase enzymes involved in the synthesis of non-ATP nucleoside triphosphates. Although it is not readily apparent as to why purine and pyrimidine metabolism would be involved in the predisposition towards virologic failure, it is possible that this could represent novel biological knowledge in this field. Currently known reasons for virologic failure include lack of adherence to drug regimen, presence of drug resistance mutations in the HIV strain, and drug interactions which might limit efficacy. In the absence of environmental heterogeneity, little is known about the etiology of virologic failure. Small sample size precludes our ability to draw conclusions about the role of nucleoside triphosphate metabolism on risk for virologic failure. Other SNP interaction models which were among the most significant results involve a SNP in the *TAP1* gene - rs735883 - and multiple SNPs in the *ABCC9* gene. *TAP1* encodes a transporter responsible for the shuttling of antigen into the endoplasmic reticulum for association with MHC class I while *ABCC9* is part of the MRP subfamily of ABC transporters associated with multi-drug resistance and codes for a protein thought to be a subunit of a pancreatic potassium channel responsible for drug-binding modulation of the channel. It could be that down-regulation of *TAP1* through mutation prevents proper immune response to the virus even after it has been affected by NNRTI action and this allows it to rebound during treatment. The results of the current study require validation with larger sample size before any firm conclusion can be drawn. The current results are meant to demonstrate the pipeline for analysis and the general approach rather than attempting to draw general statements regarding true biological associations with HIV therapy.

Despite the lack of statistical power to elucidate a significant genetic interaction, this study shows the promise of the use of Biofilter for focusing the search for gene-gene interactions during large-scale genetic association studies. The number of polymorphisms typed in association studies is nearing our limits to perform exhaustive explorations of two-way interactions during analysis. Reducing the set of interesting models to evaluate presents itself as a capable alternative. Utilizing Biofilter to provide the set of interesting models and PLATO to perform analysis has at least three advantages over traditional exhaustive gene-gene interaction analysis. First, it partially alleviates issues of multiple comparisons. Second, interpretation of results is significantly eased due to models construction. Third, the use of regression framework allows for the adjustment of the analysis taking into account important covariates. Although the use of Biofilter might not be as promising an option in cases where very little biological knowledge exists on the phenotype being analyzed, in the case of pharmacogenomics, where extensive drug metabolism networks have been elucidated, utilizing this knowledge to direct the analysis is a superior alternative, particularly when epistasis is concerned. As the search for the genetic architecture underlying complex traits such as drug pharmacokinetics continues, utilities such as the Biofilter can play an important role. Drug response is a nuanced trait and, as such, is likely to have genetic components which are monogenic as well as those that are multi-locus. Now that whole-genome sequencing technology is almost ready for wide-spread implementation, rare genetic variation is likely also to become an important component to consider for pharmacogenomic traits. Due to the nature of rare variants, the same pathway knowledge which is exploited by Biofilter to search for epistasis should be useful to group these rare variants to look for patterns predicting drug response. In summary, Biofilter is a tool which is likely to prove invaluable for the analysis of large-scale genetic association data for complex disease, especially in pharmacogenomic data where the biological knowledge is extensive.

5. Acknowledgements

This work was supported in part by the AIDS Clinical Trials Group funded by the National Institute of Allergy and Infectious Diseases (AI68636, AI38858, AI68634, AI38855), and by virology laboratory contracts (201VC001 with Vanderbilt University, 200VC001 with the University of Alabama at Birmingham, and 200VC011 with Stanford University). Grant support included AI069439, RR000095, AI54999 (DWH), AI51966, RR024996 (RMG), AI069472, AI062435 (GKR), AI64086, AI36214 (RH), GM080178 (BJG), LM010040, HG004798, AI077505, HL065962 (MDR). Clinical Research Sites that participated in ACTG protocols ACTG 384, A5095, and 5142 and collected DNA under protocol A5128, were supported by the following grants from NIH: AI69472, AI69465, AI69532, AI69556, AI27666, AI69424, AI27660, AI69432, AI69502, AI27663, AI69477, AI69494, AI46383, AI69511, AI27658, AI69428, AI69434, AI27664, AI27661, AI69484, AI38844, AI69495, AI25903, AI69474, AI69513, AI25897, AI69501, AI25859, AI69471, AI25915, AI46370, AI69472, AI46381, AI69423, AI25868, AI69439, AI46339, AI46376, AI38858-09S1, AI69447,

AI34853, AI34835, AI69415, AI69452, AI69418, AI69450, AI69467, AI32783, AI32782, AI69419, AI46386, AI69426, AI69470, AI69471, AI69503, and AI69470.

References

- [1] UNAIDS. 2008 Report on the Global AIDS epidemic. 129-158. 2008.
- [2] Granich R, Crowley S, Vitoria M, Smyth C, Kahn JG, Bennett R *et al.* Highly active antiretroviral treatment as prevention of HIV transmission: review of scientific evidence and update. *Curr Opin HIV AIDS* 2010; **5**:298-304.
- [3] Tsibris AM, Hirsch MS. Antiretroviral therapy in the clinic. *J Virol* 2010; **84**:5458-5464.
- [4] Attia S, Egger M, Muller M, Zwahlen M, Low N. Sexual transmission of HIV according to viral load and antiretroviral therapy: systematic review and meta-analysis. *AIDS* 2009; **23**:1397-1404.
- [5] Dorenbaum A, Cunningham CK, Gelber RD, Culnane M, Mofenson L, Britto P *et al.* Two-dose intrapartum/newborn nevirapine and standard antiretroviral therapy to reduce perinatal HIV transmission: a randomized trial. *JAMA* 2002; **288**:189-198.
- [6] Hetherington S, McGuirk S, Powell G, Cutrell A, Naderer O, Spreen B *et al.* Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. *Clin Ther* 2001; **23**:1603-1614.
- [7] Hawkins T. Understanding and managing the adverse effects of antiretroviral therapy. *Antiviral Res* 2010; **85**:201-209.
- [8] d'Ettorre G, Zaffiri L, Ceccarelli G, Mastroianni CM, Vullo V. The role of HIV-DNA testing in clinical practice. *New Microbiol* 2010; **33**:1-11.
- [9] Wadelius M, Chen LY, Downes K, Ghori J, Hunt S, Eriksson N *et al.* Common VKORC1 and GGCX polymorphisms associated with warfarin dose. *Pharmacogenomics J* 2005; **5**:262-270.
- [10] Oldenburg J, Bevans CG, Muller CR, Watzka M. Vitamin K epoxide reductase complex subunit 1 (VKORC1): the key protein of the vitamin K cycle. *Antioxid Redox Signal* 2006; **8**:347-353.
- [11] GlaxoSmithKline. Ziagen Medication Guide. 2009.
- [12] Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J *et al.* HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med* 2008; **358**:568-579.
- [13] Zanella F, Lorens JB, Link W. High content screening: seeing is believing. *Trends Biotechnol* 2010; **28**:237-245.
- [14] Anderson GD. Pharmacogenetics and enzyme induction/inhibition properties of antiepileptic drugs. *Neurology* 2004; **63**:S3-S8.
- [15] Iyer L, King CD, Whittington PF, Green MD, Roy SK, Tephly TR *et al.* Genetic predisposition to the metabolism of irinotecan (CPT-11). Role of uridine diphosphate glucuronosyltransferase isoform 1A1 in the glucuronidation of its active metabolite (SN-38) in human liver microsomes. *J Clin Invest* 1998; **101**:847-854.

- [16] Wilke RA, Reif DM, Moore JH. Combinatorial pharmacogenetics. *Nat Rev Drug Discov* 2005; **4**:911-918.
- [17] Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet* 2002; **11**:2463-2468.
- [18] Culverhouse R, Klein T, Shannon W. Detecting epistatic interactions contributing to quantitative traits. *Genet Epidemiol* 2004; **27**:141-152.
- [19] Moore JH, Williams SM. Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. *Bioessays* 2005; **27**:637-646.
- [20] Bateson, W, Saunders, ER, Punnett, RC, Hurst, CC. Reports to the Evolution Committee of the Royal Society, Report II. London: Harrison and Sons; 1905.
- [21] Fisher RA. The correlations between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edinb* 1918; **52**:399-433.
- [22] Bush WS, Dudek SM, Ritchie MD. Biofilter: a knowledge-integration system for the multi-locus analysis of genome-wide association studies. *Pac Symp Biocomput* 2009; 368-379.
- [23] Haas DW, Wilkinson GR, Kuritzkes DR, Richman DD, Nicotera J, Mahon LF *et al.* A multi-investigator/institutional DNA bank for AIDS-related human genetic studies: AACTG Protocol A5128. *HIV Clin Trials* 2003; **4**:287-300.
- [24] Clifford DB, Evans S, Yang Y, Acosta EP, Goodkin K, Tashima K *et al.* Impact of efavirenz on neuropsychological performance and symptoms in HIV-infected individuals. *Ann Intern Med* 2005; **143**:714-721.
- [25] Gulick RM, Ribaldo HJ, Shikuma CM, Lalama C, Schackman BR, Meyer WA, III *et al.* Three- vs four-drug antiretroviral regimens for the initial treatment of HIV-1 infection: a randomized controlled trial. *JAMA* 2006; **296**:769-781.
- [26] Gulick RM, Ribaldo HJ, Shikuma CM, Lustgarten S, Squires KE, Meyer WA, III *et al.* Triple-nucleoside regimens versus efavirenz-containing regimens for the initial treatment of HIV-1 infection. *N Engl J Med* 2004; **350**:1850-1861.
- [27] Haubrich RH, Riddler SA, DiRienzo AG, Komarow L, Powderly WG, Klingman K *et al.* Metabolic outcomes in a randomized trial of nucleoside, nonnucleoside and protease inhibitor-sparing regimens for initial HIV treatment. *AIDS* 2009; **23**:1109-1118.
- [28] Riddler SA, Jiang H, Tenorio A, Huang H, Kuritzkes DR, Acosta EP *et al.* A randomized study of antiviral medication switch at lower- versus higher-switch thresholds: AIDS Clinical Trials Group Study A5115. *Antivir Ther* 2007; **12**:531-541.
- [29] Robbins GK, De G, V, Shafer RW, Smeaton LM, Snyder SW, Pettinelli C *et al.* Comparison of sequential three-drug regimens as initial therapy for HIV-1 infection. *N Engl J Med* 2003; **349**:2293-2303.
- [30] Shafer RW, Smeaton LM, Robbins GK, De G, V, Snyder SW, D'Aquila RT *et al.* Comparison of four-drug regimens and pairs of sequential three-drug regimens as initial therapy for HIV-1 infection. *N Engl J Med* 2003; **349**:2304-2315.
- [31] Grady BJ, Torstenson ES, Dudek SM, Giles P, Ritchie MD. Finding Unique Filter Sets in PLATO: A Precursor to Efficient Interaction Analysis in GWAS Data. *Proceedings of the Pacific Symposium in Biocomputing* 2009.