

# Genetic studies in clinical trials and observational cohorts

**David W. Haas, MD** *Vanderbilt University School of Medicine, Nashville, Tennessee*

Pharmacogenomic research strives to understand the influence of human genetics on interindividual variability in response to medications. This field holds great promise to improve treatment outcomes for HIV infection and its complications. As we increase our ability to precisely define relationships between genetic polymorphisms and antiretroviral treatment response, so will the likelihood that human genetic testing will be used in clinical practice to inform antiretroviral prescribing. In addition to advancing the paradigm of “personalized medicine” through genetic testing, knowledge of genotype-phenotype associations may also help large HIV-infected populations by informing guidelines for antiretroviral therapy in resource-limited settings. For example, suppose it is discovered that human genetic variant X predicts suboptimal response to drug Y (decreased efficacy and/or increased toxicity), and genetic variant X is disproportionately represented in persons of the predominant geographic ancestry of country Z. Such information could help inform approaches to country-wide roll-out of drug Y in country Z, especially if previous experience is limited for drug Y in country Z.

## **ELICIDATING PHARMACOGENETIC ASSOCIATIONS WITH DNA REPOSITORIES**

Many associations have been reported between human genetic polymorphisms and responses to antiretroviral drugs. Such associations encompass pharmacokinetics, virologic and immunologic responses, and drug toxicities that include hypersensitivity reactions, hepatotoxicity, peripheral lipoatrophy, dyslipidemia, central nervous system intolerance, peripheral neuropathy, and hyperbilirubinemia [1-25]. Because of the risk of false discovery with association studies [26], most reported genotype-phenotype associations for HIV medications must be considered tentative until replicated in additional studies.

Available laboratory technologies for identifying genetic variability are rapidly evolving toward increased throughput and decreased cost. In the foreseeable future it may be feasible to sequence the entire human genome (approximately 3 billion base positions) for any person at reasonable cost. Despite such technological progress, deciphering relationships between genetic polymorphisms and

outcomes during HIV disease and its therapy will remain an immense computational and statistical challenge.

When clinical trials are designed or when observational cohorts are established it cannot be known a priori which genes will ultimately be most relevant. Investigators must therefore establish DNA banks for future unplanned analyses. Several groups have established banks of stored DNA to explore relationships between human genetics and responses to HIV therapy. An important repository is maintained by the NIH-funded AIDS Clinical Trials Group (ACTG). Since 2002, all ACTG study participants have had the opportunity to donate specimens to the ACTG Human DNA Repository under protocol A5128 [27]. Over 9000 different ACTG clinical trials participants have contributed specimens, accrual is ongoing, and DNA yields have been excellent. At least 17 association studies are already ongoing or have been completed using these specimens.

Other groups have established DNA repositories linked to well-characterized HIV observational cohorts, and have used these repositories to study genetic predictors of HIV treatment response. For example, data and specimens from the Western Australian HIV Cohort and the Swiss HIV Cohort Study have been used extensively to characterize important responses to HIV therapeutics. Such projects complement other well-established cohorts that have historically focused on relationships between human genetics and the natural history of untreated HIV disease, such as the NIH-funded Multicenter AIDS Cohort Study (MACS).

## **VALIDATING GENETIC ASSOCIATION STUDY RESULTS**

Genetic association study design typically involves first genotyping a “discovery” dataset to identify putative associations between genetics and phenotypes of interest (realizing that there may be many spurious associations). Promising polymorphisms are then carried forward for validation in at least one separate “replication” dataset. The size for each sequential dataset needed to minimize the risk of rejecting true associations (Type II error) depends on many factors. Among these are the underlying strength of association, the extent of

classification error for the outcome of interest, and possible confounding by non-genetic factors. Relatively small sample sizes may be sufficient to detect associations between pharmacokinetic parameters and functional variants in drug metabolism genes because such associations may be strong, the outcome variable is continuous, the likelihood of classification error is low, and confounding by non-genetic factors is usually minimal.

In contrast, treatment response (whether efficacy or toxicity) is a complex trait which depends on many factors, not just pharmacokinetics. Larger sample sizes are therefore typically required to detect associations between human genetics and treatment response. There are some exceptions, however, such as when there is high penetrance genetic susceptibility to a particular toxicity (e.g. *HLA\*B5701* and abacavir hypersensitivity reaction) or when a functional variant in a drug metabolizing gene has an extreme impact on pharmacokinetics. The sample size required for any given study will increase as the magnitude of association one hopes to detect decreases. For this reason, publications describing negative findings from genetic association studies should provide sufficient statistical information regarding the magnitude of association that could have been missed.

A great concern with genetic association studies is the risk of false discovery (Type I error). There are several strategies to reduce the number of spurious associations, the most important being replication in multiple separate studies. As genomic assay methodologies have progressed, one can now readily characterize over 500,000 distinct polymorphisms (albeit at considerable expense). In any association study of this scope there will be over 5,000 spurious associations at  $P < 0.01$ , even for traits with no genetic component. If these 5,000 “significant” polymorphisms are carried forward into a replication dataset, over 50 are still expected to remain spuriously associated at  $P < 0.01$ .

Other strategies to reduce false discovery include considering only polymorphisms for which there is strong biologic plausibility. For example, one may choose to study only genes that are involved in drug metabolism and transport, in which case the number of polymorphisms assayed will be far less than with whole genome scanning. However, for studies that use treatment response as the outcome of interest, one may miss novel associations by just studying genes involved in drug metabolism and transport. One may also reduce the likelihood of false discovery by applying statistical methodologies such as bootstrapping or permutation testing, which attempt to replicate associations across repeated subsets of a

single dataset. Yet another approach involves correcting for multiple comparisons by demanding stringent P values. Unfortunately, for association studies that assay very many polymorphisms, traditional correction approaches are not suitable since they demand extremes of statistical significance that are attainable for very few genotype-phenotype associations. This may cause inappropriate rejection of clinically relevant associations.

## CLINICAL TRIALS AND OBSERVATIONAL COHORTS ARE COMPLEMENTARY

Access to multiple complementary datasets is critical given the importance of validating initial genetic associations in independent studies, particularly for HIV pharmacogenomics. This is because antiretroviral drugs are increasingly being prescribed into diverse genetic backgrounds worldwide. Recent initiatives such as the HapMap Project are greatly facilitating genomic investigation by elucidating genomic structure and allelic frequencies among populations with different ancestral origins [28]. Allelic frequencies of polymorphisms that are important for treatment response may vary considerably among populations.

To our knowledge, every allelic variant suggested to affect antiretroviral pharmacokinetics or treatment response varies significantly in frequency among populations. For example, *CYP2B6* 516G>T is more frequent in African Americans than in European Americans [9], as is *CYP3A5* 6986A>G [11], *CYP2C19* slow metabolizer genotypes are more frequent in Asians than in other populations [29,30], and *ABCB1* position 3435 TT homozygosity is more frequent in European Americans than in African Americans [11,31]. Frequency differences among populations are also seen with target gene polymorphisms implicated in HIV treatment responses. For example *HFE* 845G>A is more frequent in European Americans than in African Americans [32], as is *HLA\*B5701* [33]. This suggests that the predictive value of human genetic testing may vary among populations.

There is no single best study population for performing human genetic association studies in the context of HIV therapeutics. Both clinical trials and observational cohorts are extremely useful for such studies. As noted in Table 1, each study type has distinct advantages for genetic association studies.

One challenge with identifying genotype-phenotype associations in population-based studies is defining the appropriate control group that shares all relevant factors except phenotype [34]. In clinical trials randomization helps achieve this goal by assuring that, on average, equal numbers of subjects with a given genotype will be assigned to different

treatments. Another strength of clinical trials is that participants may be assigned to receive highly standardized regimens. This may help to minimize confounding by non-genetic factors, and reduce the need for multivariate analysis to demonstrate associations. In addition, double blinding of study medications in clinical trials can minimize bias introduced by a priori assumptions regarding drug toxicity or efficacy. For example, in ACTG study A5095 in which abacavir was double-blinded and placebo-controlled, abacavir hypersensitivity reactions were diagnosed as frequently in abacavir placebo recipients as in individuals randomized to receive active drug [35]. Such classification error for

**Table 1.** Clinical trials and observational cohorts for genetic association studies of antiretroviral drugs

---

*Advantages of clinical trials*

- Randomization minimizes confounding by covariates
- Study regimens are highly standardized
- Drugs may be double blinded, placebo controlled
- Can generate data on drugs pre-approval
- Inclusion/exclusion criteria are rigorously defined

*Advantages of clinical cohorts*

- May more closely reflect “real world” situation
  - Flexibility to collect complete set of clinical data
  - Less likely than clinical trials to exclude patients
  - Potentially larger sample sizes and lower cost
- 

this toxicity would decrease power to detect any genetic association, and its negative predictive value, but such confounding may not have been appreciated without blinding. Early phase clinical trials of new compounds may also generate pharmacogenomic data for drugs that are not yet approved for routine clinical use, and are therefore not represented in observational cohorts. In some trials, strict inclusion and exclusion criteria will tend to assure that all individuals with a drug exposure are relatively similar, which will again reduce confounding and hence sample sizes required for an association study.

Observational cohorts also have advantages. Because virtually any HIV-infected individual may enter into a cohort, the “real world” situation may be more accurately reflected by cohorts. In addition, unlike clinical trials which for efficiency tend to collect only the pre-defined data that are needed to address specific study objectives, cohorts may have greater leeway to collect more diverse data. In fact, some cohort studies collect nearly all data that are available from routine clinical care. The broad inclusion criteria for many cohorts may also increase the generalizability of study findings, and may allow certain non-genetic (environmental) factors to be more readily incorporated into analyses. For example, inclusion of individuals at greatest risk for drug toxicity may increase the likelihood of

identifying associations, and reduce the required sample size. Finally, because costs to establish and maintain observational cohorts are far less than costs for prospective clinical trials, available sample sizes may be larger for cohorts.

### CLOSING REMARKS

In summary, associations are increasingly being reported between human genetic variants and responses to HIV therapies. However, this does not assure that such information will ever be used to improve treatment outcomes among the approximately 40 million individuals who now live with HIV. Some aspects of how medical research is conducted may impede open collaboration in this area, including pathways to academic promotion and independent funding. To make optimal progress will require a open collaboration among investigators who lead various clinical trials groups and observational cohorts. Only in this way will pharmacogenomics have the greatest chance to benefit persons living with HIV.

### REFERENCES

- (1) Lillibridge JH, Lee CA, Pithavala YK. The role of polymorphic CYP2C19 in the metabolism of nelfinavir mesylate. Presented at XII Annual Meeting and Exposition of the American Association of Pharmaceutical Scientists; San Francisco American Association of Pharmaceutical Scientists, Arlington, VA. November 1998, abstract 3035.
- (2) Zucker SD, Qin X, Rouster SD, et al. Mechanism of indinavir-induced hyperbilirubinemia. *Proc Natl Acad Sci USA* 2001;98:12671-6.
- (3) Mallal S, Nolan D, Witt C, et al. Association between presence of HLA-B\*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* 2002;359:727-32.
- (4) Hetherington S, Hughes AR, Mosteller M, et al. Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. *Lancet* 2002;359:1121-2.
- (5) Fellay J, Marzolini C, Meaden ER, et al. Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenetics study. *Lancet* 2002;359:30-6.
- (6) Maher B, Alfirevic A, Vilar FJ, Wilkins EG, Park BK, Pirmohamed M. TNF-alpha promoter region gene polymorphisms in HIV-positive patients with lipodystrophy. *AIDS* 2002;16:2013-8.
- (7) O'Mara et al. Population pharmacodynamic assessment of atazanavir exposure, uridine diphosphate-glucuronosyl transferase (UGT) 1A1 genotype and safety in healthy subjects. 41st Interscience Conference on Antimicrobial Agents and Chemotherapy, Chicago, IL. 2003, abstract 3051.
- (8) Nolan D, Moore C, Castley A, et al. Tumour necrosis factor-alpha gene -238G/A promoter polymorphism associated with a more rapid onset of lipodystrophy. *AIDS* 2003;17:121-3.

- (9) Haas DW, Ribaldo HJ, Kim RB, et al. Pharmacogenetics of efavirenz and central nervous system side effects: an Adult AIDS Clinical Trials Group study. *AIDS* 2004;18:2391-400.
- (10) Price P, James I, Fernandez S, French MA. Alleles of the gene encoding IL-1alpha may predict control of plasma viraemia in HIV-1 patients on highly active antiretroviral therapy. *AIDS* 2004;18:1495-501.
- (11) Haas DW, Smeaton LM, Shafer RW, et al. Pharmacogenetics of Long-Term Responses to Antiretroviral Regimens Containing Efavirenz and/or Nelfinavir: An Adult AIDS Clinical Trials Group Study. *J Infect Dis* 2005;192:1931-42.
- (12) Rotger M, Colombo S, Furrer H, et al. Influence of CYP2B6 polymorphism on plasma and intracellular concentrations and toxicity of efavirenz and nevirapine in HIV-infected patients. *Pharmacogenetics Genomics* 2005;15:1-5.
- (13) Mouly SJ, Matheny C, Paine MF, et al. Variation in oral clearance of saquinavir is predicted by CYP3A5\*1 genotype but not by enterocyte content of cytochrome P450 3A5. *Clin Pharmacol Ther* 2005;78:605-18.
- (14) Hulgán T, Haas DW, Haines JL, et al. Mitochondrial haplogroups and peripheral neuropathy during antiretroviral therapy: an adult AIDS clinical trials group study. *AIDS* 2005;19:1341-9.
- (15) Martin AM, Nolan D, James I, et al. Predisposition to nevirapine hypersensitivity associated with HLA-DRB1\*0101 and abrogated by low CD4 T-cell counts. *AIDS* 2005;19:97-9.
- (16) Tarr PE, Taffe P, Bleiber G, et al. Modeling the influence of APOC3, APOE, and TNF polymorphisms on the risk of antiretroviral therapy-associated lipid disorders. *J Infect Dis* 2005;191:1419-26.
- (17) Ribaldo HJ, Haas DW, Tierney C, et al. Pharmacogenetics of plasma efavirenz exposure after treatment discontinuation: an Adult AIDS Clinical Trials Group Study. *Clin Infect Dis* 2006;42:401-7.
- (18) Haas DW, Bartlett JA, Andersen JW, et al. Pharmacogenetics of nevirapine-associated hepatotoxicity: an Adult AIDS Clinical Trials Group collaboration. *Clin Infect Dis* 2006;43:783-6.
- (19) Ritchie MD, Haas DW, Motsinger AA, et al. Drug transporter and metabolizing enzyme gene variants and nonnucleoside reverse-transcriptase inhibitor hepatotoxicity. *Clin Infect Dis* 2006;43:779-82.
- (20) Rodriguez NS, Barreiro P, Rendon A, et al. Plasma levels of atazanavir and the risk of hyperbilirubinemia are predicted by the 3435C-->T polymorphism at the multidrug resistance gene 1. *Clin Infect Dis* 2006;42:291-5.
- (21) Hulgán T, Haas D, Haines J, et al. Non-synonymous Mitochondrial DNA Polymorphisms and Peripheral Neuropathy during NRTI Therapy in ACTG Study 384. Presented at 13th Conference on Retroviruses and Opportunistic Infections, Denver, CO. February 2006, abstract 350.
- (22) Foulkes AS, Wohl DA, Frank I, et al. Associations among race/ethnicity, ApoC-III genotypes, and lipids in HIV-1-infected individuals on antiretroviral therapy. *PLoS Med* 2006;3:e52.
- (23) Ranade K, Ploughman L, Parker R, et al. A Single Nucleotide Polymorphism in the Resistin Gene Is Associated with Adverse Metabolic Changes on HAART: An Exploratory Pharmacogenetic Association Study of A5005s, the Metabolic Sub-study of ACTG 384. Presented at 13th Conference on Retroviruses and Opportunistic Infections, Denver, CO. February 2006, abstract 763.
- (24) Kallianpur AR, Hulgán T, Canter J, et al. Hemochromatosis (HFE) gene mutations and peripheral neuropathy during antiretroviral therapy. *AIDS* 2006;20:1503-13.
- (25) Haas DW, Geraghty DE, Andersen J, et al. Immunogenetics of CD4 Lymphocyte Count Recovery during Antiretroviral Therapy: An AIDS Clinical Trials Group Study. *J Infect Dis* 2006;194:1098-107.
- (26) Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet* 2001;29:306-9.
- (27) 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;4/5:287-300.
- (28) Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. *Nature* 2005;437:1299-320.
- (29) De Morais SM, Wilkinson GR, Blaisdell J, Meyer UA, Nakamura K, Goldstein JA. Identification of a new genetic defect responsible for the polymorphism of (S)-mephenytoin metabolism in Japanese. *Molecular Pharmacol* 1994;46:594-8.
- (30) De Morais SM, Wilkinson GR, Blaisdell J, Nakamura K, Meyer UA, Goldstein JA. The major genetic defect responsible for the polymorphism of (S)-mephenytoin metabolism in humans. *J Biol Chem* 1994;269:15419-22.
- (31) Kim RB, Leake BF, Choo EF, et al. Identification of functionally variant MDR1 alleles among European Americans and African Americans. *Clin Pharmacol Ther* 2001;70:189-99.
- (32) Steinberg KK, Cogswell ME, Chang JC, et al. Prevalence of C282Y and H63D mutations in the hemochromatosis (HFE) gene in the United States. *JAMA* 2001;285:2216-22.
- (33) Nolan D, Gaudieri S, Mallal S. Pharmacogenetics: a practical role in predicting antiretroviral drug toxicity? *J HIV Ther* 2003;8:36-41.
- (34) Merikangas KR, Swendsen JD. Genetic epidemiology of psychiatric disorders. *Epidemiol Rev* 1997;19:144-55.
- (35) Gulick RM, Ribaldo HJ, Shikuma CM, et al. Three- vs four-drug antiretroviral regimens for the initial treatment of HIV-1 infection: a randomized controlled trial. *JAMA* 2006;296:769-81.