2013

Host Genomic Influences on HIV/AIDS

Stephen J. O'Brien
St. Petersburg State University - Russia, sobrien1@nova.edu

Sher L. Hendrickson
Shepherd University

Follow this and additional works at: http://nsuworks.nova.edu/cnso_bio_facarticles

Part of the Genetics and Genomics Commons, Immunology and Infectious Disease Commons, and the Medicine and Health Sciences Commons

NSUWorks Citation
http://nsuworks.nova.edu/cnso_bio_facarticles/745

This Article is brought to you for free and open access by the Department of Biological Sciences at NSUWorks. It has been accepted for inclusion in Biology Faculty Articles by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.
Introduction
Thirty-one years have passed since reports of severely immune-compromised gay men in San Francisco and New York first drew attention to the AIDS epidemic [1,2]. It took 3 years to identify the cause as human immunodeficiency virus (HIV)-1, a lentivirus that subverts the lymphoid system to become viral factories producing 1-2 billion virions per day. HIV-1 infects and destroys CD4-bearing lymphocytes, decreasing the victim’s ability to overcome normally innocuous agents such as Pneumocystis, Candida, cytomegalovirus and a score of other opportunistic infections (Box 1). Left untreated, these infections and certain rare cancers contribute further to the immune system collapse and to the demise of the patients. An effective vaccine is yet to be developed, and anti-retroviral drugs are not ideal since they do not work well for all patients, can generate adverse side effects and do not eliminate the virus from an infected patient. Since the first reports of AIDS in 1981, the world has seen more than 60 million cases of HIV infections, and 30 million victims have died as a consequence, 90% of them in the developing world [3].

The genomic era began eleven years ago with the release of the draft sequence of the human genome. A complete human gene set and DNA sequence was available to help combat the more inscrutable deadly chronic and infectious diseases [4], and this has enabled the investigation of HIV/AIDS using the new tools of genomics. AIDS exhibits abundant epidemiological heterogeneity, much of which could be attributed to host genetic factors. Beginning in the early 1980s, epidemiologists began collecting longitudinal cohorts of at-risk AIDS populations to describe that heterogeneity. Many collaborated with geneticists, who used population genetics-based association analyses to uncover genes with natural variants that exerted an effect on HIV infection, on the dynamics of AIDS progression, and on the results of highly active anti-retroviral therapy (HAART). This review will attempt to update the status of discovered host genetic influences on stages of AIDS beyond the informative reviews that have appeared to date [5-10]. We will highlight those host gene influences that are credible, replicated and implicated in AIDS disease, and we will discuss their clinical relevance in stemming the spread of HIV.

Candidate AIDS restriction genes: the case of CCR5-Δ32
Before the first of ten HIV-AIDS genome-wide association studies (GWAS) appeared [11], single nucleotide polymorphism (SNP) variants in numerous candidate genes were suggested as being associated with HIV or AIDS in patients. Several hundred candidate gene SNPs were tested for association with AIDS progression and for HIV transmission. Over 35 plausible AIDS restriction genes (ARGs) have been identified (Table 1). Their discovery, replication and functional interpretation have been reviewed in depth, and these reviews should be consulted to appreciate the details uncovered around each ARG discovery [5-10].

The first and perhaps most provocative ARG involved the description of CCR5-Δ32, a 32 base pair deletion frameshift mutation that truncates C-C chemokine receptor 5 (CCR5), the HIV entry receptor on lymphoid cells [12-14] (see Box 2 for a discussion of the origin of this mutation). Epidemiological studies showed that individuals homozygous for CCR5-Δ32 had a 100-fold reduction in HIV infection incidence: the genotype seems to confer near complete protection from HIV infection (Figure 1). Individuals in large cohort studies that are heterozygous for CCR5-Δ32, although susceptible to infection, consistently have a delayed onset of AIDS by two to four years, likely due to the diminishing onset of
Further, AIDS patients on HAART live longer when they are heterozygous for the CCR5-Δ32 mutation, due to faster HIV viral load suppression, and progress to AIDS more slowly than CCR5+/+ patients on HAART (Figure 1c) [16]. It is clear that latent HIV continues its pathogenesis in these patients, even in the presence of effective anti-retroviral treatment. We therefore need to develop better AIDS therapies.

**CCR5 in clinical practice**

Individuals homozygous for CCR5-Δ32 seem to be relatively healthy into advanced age, suggesting that
### Table 1. Candidate AIDS restriction genes

<table>
<thead>
<tr>
<th>Year</th>
<th>Gene symbol</th>
<th>Allele</th>
<th>Mode</th>
<th>Effect</th>
<th>Citation(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>CCR5</td>
<td>Δ32</td>
<td>Recessive</td>
<td>Prevents infection</td>
<td>[12-14]</td>
</tr>
<tr>
<td>1996</td>
<td>CCR5</td>
<td>Δ32</td>
<td>Dominant</td>
<td>Delays AIDS</td>
<td>[12]</td>
</tr>
<tr>
<td>1997</td>
<td>CCR2</td>
<td>64I</td>
<td>Dominant</td>
<td>Delays AIDS</td>
<td>[80]</td>
</tr>
<tr>
<td>1998</td>
<td>CCR5P</td>
<td>P1</td>
<td>Recessive</td>
<td>Accelerates AIDS</td>
<td>[81]</td>
</tr>
<tr>
<td>1998</td>
<td>SDF1</td>
<td>3'A</td>
<td>Recessive</td>
<td>Delays AIDS</td>
<td>[82]</td>
</tr>
<tr>
<td>1999</td>
<td>CCR5</td>
<td>Δ32</td>
<td>Dominant</td>
<td>Prevents lymphoma</td>
<td>[15]</td>
</tr>
<tr>
<td>1999</td>
<td>HLA</td>
<td>A,B,C'</td>
<td>Co-dominant</td>
<td>Accelerates AIDS</td>
<td>[83]</td>
</tr>
<tr>
<td>2000</td>
<td>IL10</td>
<td>5'A</td>
<td>Dominant</td>
<td>Limits infection</td>
<td>[84]</td>
</tr>
<tr>
<td>2001</td>
<td>IL10</td>
<td>5'A</td>
<td>Dominant</td>
<td>Accelerates AIDS</td>
<td>[84]</td>
</tr>
<tr>
<td>2002</td>
<td>RANTES</td>
<td>−403A</td>
<td>Co-dominant</td>
<td>Accelerates AIDS</td>
<td>[86]</td>
</tr>
<tr>
<td>2002</td>
<td>KIR</td>
<td>3DS1</td>
<td>Epistatic</td>
<td>Delays AIDS</td>
<td>[28,29]</td>
</tr>
<tr>
<td>2003</td>
<td>EOTAXIN-MCP1</td>
<td>Hap7</td>
<td>Dominant</td>
<td>Enhances infection</td>
<td>[87]</td>
</tr>
<tr>
<td>2003</td>
<td>HLA</td>
<td>B*57</td>
<td>Co-dominant</td>
<td>Delays AIDS</td>
<td>[88]</td>
</tr>
<tr>
<td>2003</td>
<td>IFNG</td>
<td>179T</td>
<td>Dominant</td>
<td>Accelerates AIDS</td>
<td>[89]</td>
</tr>
<tr>
<td>2003</td>
<td>CXC6</td>
<td>E3K</td>
<td>Dominant</td>
<td>Accelerates PCP</td>
<td>[90]</td>
</tr>
<tr>
<td>2004</td>
<td>APOBEC3G</td>
<td>H186R</td>
<td>Recessive</td>
<td>Accelerates AIDS</td>
<td>[91]</td>
</tr>
<tr>
<td>2004</td>
<td>DCSIGN</td>
<td>−336T</td>
<td>Dominant</td>
<td>Decreases infection</td>
<td>[92]</td>
</tr>
<tr>
<td>2006</td>
<td>HLA</td>
<td>B27</td>
<td>Co-dominant</td>
<td>Delays AIDS</td>
<td>[93]</td>
</tr>
<tr>
<td>2006</td>
<td>TSG101</td>
<td>Hap2</td>
<td>Dominant</td>
<td>Accelerates AIDS</td>
<td>[94]</td>
</tr>
<tr>
<td>2006</td>
<td>TRIMS</td>
<td>Hap4</td>
<td>Dominant</td>
<td>Increases infection</td>
<td>[95]</td>
</tr>
<tr>
<td>2007</td>
<td>CUL5</td>
<td>Hapi</td>
<td>Co-dominant</td>
<td>Accelerates CD4 loss</td>
<td>[96]</td>
</tr>
<tr>
<td>2007</td>
<td>PPIA (cyclophilin A)</td>
<td>SNP-4</td>
<td>Dominant</td>
<td>Accelerates AIDS</td>
<td>[97]</td>
</tr>
<tr>
<td>2007</td>
<td>HLA</td>
<td>Bv4</td>
<td>Dominant</td>
<td>Reduces HIV Transmission</td>
<td>[98]</td>
</tr>
<tr>
<td>2008</td>
<td>MYH9</td>
<td></td>
<td></td>
<td>End Stage Renal Disease</td>
<td>[31]</td>
</tr>
<tr>
<td>2008</td>
<td>MYH9</td>
<td></td>
<td></td>
<td>HIV FSGN</td>
<td>[30]</td>
</tr>
<tr>
<td>2008</td>
<td>mtDNA</td>
<td>Hap-J, USa</td>
<td>Dominant</td>
<td>Accelerates AIDS</td>
<td>[56]</td>
</tr>
<tr>
<td>2008</td>
<td>mtDNA</td>
<td>Hap-H</td>
<td>Dominant</td>
<td>Increases</td>
<td>[55]</td>
</tr>
<tr>
<td>2008</td>
<td>mtDNA</td>
<td>Hap-J</td>
<td>Dominant</td>
<td>Delays CMV-NRD</td>
<td>[99]</td>
</tr>
<tr>
<td>2009</td>
<td>HCPS</td>
<td>T&gt;G; rs2395029</td>
<td>Dominant</td>
<td>HIV set point</td>
<td>[33]</td>
</tr>
<tr>
<td>2009</td>
<td>HLA</td>
<td>rs9264942</td>
<td>Dominant</td>
<td>HIV set point</td>
<td>[33]</td>
</tr>
<tr>
<td>2009</td>
<td>PROX1</td>
<td>Hap-CGT</td>
<td>Recessive</td>
<td>Delays AIDS progression</td>
<td>[37]</td>
</tr>
<tr>
<td>2009</td>
<td>APOBEC3B</td>
<td>ΔIV</td>
<td>Recessive</td>
<td>Increases infection</td>
<td>[100]</td>
</tr>
<tr>
<td>2010</td>
<td>PECI</td>
<td>G</td>
<td>Dominant</td>
<td>Accelerates AIDS</td>
<td>[52]</td>
</tr>
<tr>
<td>2010</td>
<td>ACSM4</td>
<td>A</td>
<td>Co-dominant</td>
<td>Delays AIDS</td>
<td>[52]</td>
</tr>
<tr>
<td>2010</td>
<td>NCOR2</td>
<td>T</td>
<td>Dominant</td>
<td>Increases infection</td>
<td>[57]</td>
</tr>
<tr>
<td>2010</td>
<td>IDH1</td>
<td>C</td>
<td>Dominant</td>
<td>Prevents infection</td>
<td>[57]</td>
</tr>
</tbody>
</table>

All these genes were discovered or validated using the NCI-Laboratory of Genomic Diversity. Eight AIDS cohorts comprising more than 10,000 study participants were used in these studies from 1985 to 2011 [5,10,79,101].
CCR5 function is dispensable in the human genome, a notion supported by the redundancy of other human chemokine receptors that also interact with specific chemokine ligands of CCR5 [17]. The knowledge that CCR5 is required for HIV-1 infection of lymphoid cells in exposed people, plus the suspicion that CCR5 function is dispensable for good health [18] (see Box 3), stimulated pharmaceutical development of a dozen new anti-AIDS treatments termed HIV entry inhibitors. Entry inhibitors were designed to interfere with HIV–CCR5 cell receptor binding as a block to HIV pathogenesis [19-21]. Two powerful HIV entry inhibitors, enfuvirtide (T20) and maraviroc, were approved for AIDS patient treatment by the Food and Drug Administration (FDA) in 2003 and 2007, respectively. The rapid development of this new avenue for AIDS therapy was clearly encouraged by the characterization of HIV resistance in CCR5-Δ32 homozygotes (the 30-plus other HAART drugs all target HIV-encoded enzymes: reverse transcriptase – targeted by both nucleotide reverse transcriptase inhibitors and non-nucleoside reverse transcriptase inhibitors – and the HIV-encoded protease).

A remarkable example of CCR5-Δ32 clinical translation to the bedside involves the treatment of Timothy Leigh Brown, dubbed the ‘Berlin patient’ [22]. Brown was a 25-year-old American living in Berlin in 1995 when he was diagnosed with HIV infection. Despite receiving sporadic anti-HIV medication, his condition deteriorated and he was diagnosed with acute myelogenous leukemia in 2006. His AIDS clinician, Gero Hutter, prescribed a stem cell transplant of healthy HIV-negative donor lymphocytes following x-irradiation of his lymphoid system [23]. Hutter searched the European bone marrow databases to find an HLA-matched donor who was homozygous for CCR5-Δ32, so as to block HIV transmission to the donor graft cells. After the stem cell transplant in February 2007, Brown’s anti-retroviral treatment was suspended to allow the graft to grow, though monthly blood tests were initiated to detect the inevitable latent HIV rebound. But rebound never occurred. Three years later, Hutter and his team reported that Brown had consistently tested negative for circulating plasma HIV since the transplant [23]. Subsequent rather invasive biopsies of multiple tissues (plasma, gut, brain, testis, liver) using hyper-sensitive virus assays were also negative, indicating that his body had been cleared of replicating HIV, the first such event recorded in history [22-24]. Though no one is certain that each and every one of the scores of human tissues that can become infected with HIV are clear in Mr Brown's body, as of now – five years later – his case is being heralded as the single documented clearance or apparent cure of HIV infection.

Bone marrow and stem cell transplants are of limited promise as a universal therapy. They are expensive (circa US$250,000) and not very safe, primarily because even HLA-matched donors are immunologically allogeneic (that is, non-self). The immune system of the recipient must be destroyed, deadly graft versus host reaction against the host is common, and a dangerous transplantation therapy must be followed by a lifelong...
regimen of anti-rejection therapies. A promising alternative to this may be immunologically syngeneic (self) grafts with CCR5 knockout strategies. Carl June and Bruce Levine at University of Pennsylvania have developed such a protocol using zinc finger targeted knockout of normal CCR5 genes in HIV-infected patients [25]. Twelve such patients have been infused with their own-engineered CCR5–/– lymphocytes. The results are promising to date, with donor cell reconstitution observed and few safety issues encountered, and several patients remained plasma HIV-negative several months post transplant infusion.

News of the Berlin patient’s amazing recovery travelled quickly in the AIDS research community and beyond. Attention was also paid to the fascinating secondary observation that the Berlin patient has never developed graft versus host disease (GVHD), a deadly reaction that afflicts 30–50% of bone marrow and stem transplant recipients. Transplant researchers at University of Pennsylvania reasoned that perhaps the CCR5-Δ32 genotype of the Berlin patient’s transplant donor might have prevented his GVHD, this would affirm a role for intact CCR5 in GVHD. The availability of safe and FDA approved CCR5 antagonist drugs for AIDS, such as maraviroc, prompted a Phase I and II clinical trial to see if maraviroc-treated transplant patients might avoid GVHD, as with the Berlin patient [26]. A group of 38 cancer patients treated with maraviroc post-transplant demonstrated a dramatic reduction in GVHD without any major adverse side effects. If confirmed by a larger Phase III trial, CCR5 may therefore be implicated as a key player in cancer treatments as well as AIDS therapies. It should be noted that maraviroc is considered to be a safe drug and that CCR5-Δ32/32 homozygotes are generally healthy. If such bold new therapies do prove efficacious, they will stand upon the findings of the CCR5-Δ32/Δ32 homozygotes’ genetic resistance to HIV acquisition and the apparent success with the Berlin patient.
AIDS resistance genes beyond CCR5

Each of the genes listed in Table 1 has the potential to have a detailed narrative such as that described for HLA-B*5701, which confers a complete loss of function on CCR5. For example, a rare HLA allele in the population can protect individuals from HIV infection. The HLA-B*5701 allele is known to prevent the entry of HIV into CD4+ T cells by binding to the CCR5 receptor. This interaction is critical for HIV replication, and the presence of HLA-B*5701 can prevent this interaction, thereby reducing HIV infection risk.

SNP variants within loci that exert powerful anti-HIV restrictions in vitro and against other viruses (TRIM5, APOBEC, TSG101, CIITA) have shown modest effects on AIDS progression in association studies (Table 1). Unlike CCR5-Δ32, which confers a complete loss of function mutation, the consequence of most other ARG SNP variants are limited by quantitative influences on gene action. An important predictive influence on HIV-associated nephropathy, particularly among African Americans, appears to be associated with multi-SNP haplotypes within the MYH9 (non-muscle myosin IIA) gene [30,31]. The risk alleles are quite common in African Americans but rare in European Caucasians, which explains a 20-fold higher risk for HIV-associated nephropathy in African Americans. Taken together, combining functional data that connects ARG variants directly to HIV pathology, robust statistical gene SNP associations and clinical translation has improved our understanding of diagnosis, pathogenesis, treatment and ablation of AIDS.

HIV/AIDS GWAS

Unlike association studies of candidate genes, GWAS scan the entire genome with hundreds of thousands (if not millions) of common human SNPs to discover genome regions that statistically associate with explicit AIDS endpoints in large cohort studies (Box 1). The first GWAS for HIV progression was published in 2007 by the European CHAVI cohort that were genotyped with an Illumina 550,000 SNP array. The GWAS identified several genome-wide significant SNPs, including CCR5-Δ32 homozygotes, which was protective against AIDS, and also the object of historic selective events likely from infectious agents other than HIV in the distant past, has stimulated a search for other viruses that may co-opt CCR5 molecules as entry portals to human cells. One fascinating study using West Nile Virus (WNV) raises some troubling issues [77,78]. WNV, a mosquito-borne flavivirus that was first introduced into New York City in 1999, induces a fatal encephalitis in 20% of human infections. The CR5Δ32 homozygotes were elevated 4- to 13-fold among the survivors [78]. Clearly the intact CCR5 gene function exerts a critical defense in the face of WNV. The alternative AIDS resistance and WNV hypersensitivity poses a mixed blessing for homozygous carriers of CCR5-Δ32 and a stark reminder that genetic adaptation can come with a cost in a changing microbial environment.
the causal or operative SNP variant; (3) copy number variants can be resolved by SNP genotyping arrays; and (4) subsets of gene families can be explored using SNP array data (see below). The disadvantages of GWAS are that there are no a priori assumptions about a possible connection to AIDS as there are for candidate genes; therefore, SNP association discovery must be complemented with additional function-based implication.

The large number of SNPs interrogated requires statistical correction for multiple testing, and to avoid false positives the significance threshold must be stringent. Yet mixed among the below-threshold SNPs are a few gems, such as those in CCR5 and KIR, whose presence in pathways associated with AIDS pathogenesis make them too good to miss (for instance, many ARGs in Table 1 are not revealed by GWAS). Finally, GWAS cannot resolve rare causal variants (f <5%). Goldstein has suggested that rare variants may account for undiscovered genetic influences on AIDS and other complex diseases [42,43], but these genetic variants would not be uncovered by GWAS. Although it is impossible to predict future discoveries with confidence, the frequent occurrence of deadly infectious disease episodes in human history has clearly provided abundant selective pressures (possibly raising allele frequencies for CCR5-Δ32, HB-S, DUFy, HLA and others) elevating many disease immune defense alleles above the ‘rare allele’ frequency. We also suggest that the study design of some of the previous AIDS GWAS might be improved by assessing clinical endpoint/phenotypes less variable than viral load, set point and CD4 cell trajectory. Although viral load/set point are valid predictors of AIDS progression on a course scale, in our cohorts the signal suffers somewhat from noise, and is also less continuous than survival analyses (using Cox proportional hazards analyses for detecting ARGs that influence AIDS progression). Better approaches for resolving the true association signals that fall below genome-wide significance threshold are needed if we are to maximize the potential of AIDS GWAS, and will likely require the evaluation of gene associations from a variety of perspectives.

One recent example of a potentially ‘better approach’ is the recent GWAS that associated several SNPs within the large 1.1 Mbp PARD3B gene (chr2), a locus that regulates cell polarity in epithelial cells [40,44]. Nine of the 15 lowest P-values in the GWAS fell in a short 5 kb LD block across a single PARD3B exon. One of the nine SNPs reached genome-wide significance (P = 3.4 × 10⁻⁵), while the others fell just below that threshold. The basis for the association (Figure 2) demonstrates that the rare genotype (GG) for the SNP almost never progress to AIDS, while people that are heterozygous for this SNP progress markedly slower than those homozygous for the common allele (CC). Support for the gene association came from six angles: (1) a nearly identical effect was observed among two distinct risk groups (hemophiliacs and gay men; Figure 2b,c); (2) the SNP association result was replicated, albeit with a weaker statistical strength (P = 0.03), in independent European cohorts; (3) one of the associated SNPs is a predicted exon splicing enhancer; (4) a quantitative measure of PARD3B splice products in lymphoblastoid B-cell lines with alternative PARD3B genotypes revealed a dominant 30 to 40% reduction in PARD3B isoform transcripts; (5) in the GWAS, PARD3B alleles explained 4.3% of the epidemiological variance of AIDS progression, higher than any of the other candidate genes in Table 1 similarly assessed in the same cohorts; (6) PARD3B gene products are known to interact with SMAD proteins, signal transduction modulators that mediate gene signaling, including through a direct interaction with HIV regulatory proteins [45]. Taken together, the cumulative replication, HIV1 regulatory connection, transcription influence and association data on PARD3B offers a plausible new genetic influence of AIDS progression.

Beyond GWAS: new approaches

The PARD3B study was intentionally limited to a single endpoint or phenotype (time of progression as defined by the AIDS 1987 Centers for Disease Control) to avoid statistical penalties for multiple tests. Yet many different gene association tests are possible with the rich clinical and genetic data we have collected across the years on these retrospective AIDS cohorts; indeed, many of these tests were performed and considered supportive in the previous candidate gene studies [7,10] (Table 1). In spite of the aversion to multiple test statistical correction penalties, there is information that can be gleaned from detailed clinical data that should be explored. Hutcheson et al. [7] described useful heat plots that allow one to inspect the pattern of genetic association across many tests, as well as across small regions of the genome in strong LD. The combination of multiple non-independent statistical signals is a challenge to interpret, but in cases of known genes, such as CCR5-Δ32, HLA and IL10, the patterns are illuminating. The heat plots also provide a derivative approach to public data sharing of GWAS results without the problem of violating the informed consent and privacy constraints set out by Institutional Review Boards, which often restrict open access to human cohort data [46]. For example, posting an unabridged table of GWAS results online (odds ratios, P-values and confidence intervals), together with two-dimensional and three-dimensional visualization heat plots, would allow researchers to freely inspect the results of a published GWAS [40], in any genomic region they may discover in a different cohort, for independent replication purposes – without viewing patients’ confidential clinical or genotypic information.
Now that ten AIDS GWAS have been completed [11,30-41], it is important to share and compare these data to maximize the discoveries and minimize Type 1 (false positive) errors. Meta-analyses such as the International HIV Acquisition Consortium and the HIV Controllers Consortium are important advances that should be encouraged [35,47]. Further, most of the published studies with any statistical power involve European Caucasian patients. Newer cohorts in Africa and Asia would increase the chance of discovering adaptive variants that may have arisen in those ethnicities in recent centuries. Such cohort developments are ongoing with the CHAVI consortium [48], the Botswana Harvard Partnership [49] and the Chinese AIDS collaboration [50]. It is also important to expand interrogation of critical stages of HIV/AIDS by exome and whole genome sequence analyses to capture SNP variants that are not well represented in available genotyping arrays [51]. Whole genome sequencing will offer an unabridged assessment of gene variants in these cohorts and will likely reveal additional host factors that regulate AIDS pathogenesis.

A compromise approach between the specificity of candidate gene screens and the stringency of GWAS statistical thresholds is to explore certain candidate gene subsets represented within GWAS. This strategy limits the considered SNPs to specific subsets of candidate genes and lowers the multiple SNP test threshold for genome-wide significance. Hendrickson et al. [52] used this approach to explore the association of a group of human genes that are involved in mitochondrial metabolism, named nuclear encoded mitochondrial proteins (NEMPs) [53]. Mitochondrial metabolism is relevant because AIDS pathogenesis clearly involves mitochondrial deterioration and cell death, because mitochondrial toxicity has been linked to several adverse events of HAART, and because mitochondrial haplotypes appear to be associated with differential survival to AIDS in longitudinal cohorts [54-56].

Two tested NEMP-encoding genes displayed significant (corrected for all SNPs considered) genotypic influences
on AIDS progression: peroxisomal D3,D2-enoyl-CoA isomerase (PECI) on chromosome 6 and acyl-CoA synthetase medium-chain family member 4 (ACSM4) on chromosome 12 (Table 1). The gene subset approach allowed the detection of gene associations that would have been discounted in an agnostic full GWAS since their \( P \)-values of association significance \( (P = 6.5 \times 10^{-6} \) and \( 3.6 \times 10^{-6} \) for PECI and ACSM4, respectively) fall below a full GWAS \( P \)-value threshold.

Another GWAS subset explored was that of genes required for HIV infection of cultured human cells, termed HIV dependency factors (HDFs) [57]. In 2008, four siRNA-based in vitro studies were published that collectively identified several hundred HDF genes [58-61]. Chinn et al. [57] identified 6,380 SNPs that occurred within 278 HDF genes and used a subset of the PARDS3B GWAS study data [40] to test them for gene associations with HIV infection, AIDS progression and AIDS-defining sequelae. SNP variants in two genes, NCOR2 (also known as SMRT) and IDH1, showed significant associations with HIV infection, consistent with their selection from HDF genes. IDH1 was particularly notable, first since its HDF status was one of a few replicated in independent HDF studies [58,59], and second because IDH1 can function as both a tumor suppressor (loss of isocitrate dehydrogenase 1 (IDH1) function) and an oncogene in human glioblastomas and other neoplasms [62-64].

The zinc finger ribbon domain 1 (ZNRD1) gene (located within the HLA locus) was one of the most strongly associated variants with HIV set point and disease progression in several GWAS [11,33]. ZNRD1 also appeared as a candidate gene in one HDF screen [58], while Chinn et al. [57] implicated this gene as being associated with both HIV acquisition and HIV progression. These multiple indications of AIDS influence using different approaches (GWAS, set point and survival analyses) together make a strong case for the influence of the ZNRD1 genotype on HIV/AIDS.

**Conclusions**

Thirty years of genomics investigation of AIDS has produced some important findings that have improved understanding of the disease and informed attempts to stem its deadly pace. Many host genes that affect AIDS have been postulated through candidate gene associations, GWAS, HDF analyses, molecular virology and other approaches. Many, but not all, of the implicated genes have been independently replicated by various approaches designed to overcome the statistical challenge posed by multiple tests (for example, CCR5-Δ32, HLA, KIR, IDH1, ZFNR1 and others). Most important, at least five discoveries have led to bedside translation: (1) development of CCR5-Δ32-based HIV entry inhibitor therapy [20,21]; (2) HLA-based pharmacogenomics screening for B*5701-mediated abacavir adverse events [27]; (3) clearing of HIV from the Berlin patient by a stem cell transplant from a CCR5-Δ32/Δ32 homozygous donor led to gene therapy trials of autologous CCR5-Δ32 knockout in AIDS patients [22-24]; (4) use of the CCR5 antagonist, maraviroc, as GVHD prophylaxis in cancer patients receiving marrow and stem cell transplants [26]; and (5) IDH1 discoveries have stimulated the search for drugs that block the accumulation of 2-hydroxyglutarate for cancer [62-64]. These cancer-targeted pharmaceuticals may well equally inform HIV/AIDS therapies in the future.

The disappointments of the ARG searches are several. The total number of ARGs described account for less than 10 to 15% of the variance in AIDS progression and far less for other AIDS-associated measures: namely HIV acquisition, HAART outcomes or specific AIDS-defining conditions. This means that most of the determinants (whether genetic or non-genetic) are yet to be determined. The stringent statistical threshold for GWAS may force us to discount real variants that do not pass the genome-wide significance level; we need to develop a solution to this conundrum. In addition, genotyping arrays used to date, which contain between 500,000 and 1,000,000 SNPs, cover at best 10 to 15% of the variants in genes and gene regions that we want to interrogate, so we depend upon LD to detect the causal variants through inference. We are missing much of the common variation, especially that found outside of genes (such as in the recently described 4 million regulatory regions [65]), and nearly 100% of the rare variants. Finally, because present GWAS largely target European Caucasian ethnic groups, variants accumulated in other ethnicities have not been thoroughly screened.

The next generation of ARG GWAS will address and remedy these shortcomings as follows. Combined meta-analyses, such as the International HIV Acquisition Consortium [35,47], will help affirm the statistically uncertain discoveries that are robust. Subset analyses such as the HDF and NEMP groups described in this review will help here. In addition, better strategies for data sharing of large GWAS and association studies will go a long way to encourage independent inspection, replication and composite interpretations. Bigger genotyping arrays (for example, Illumina 2.5M Infinium arrays) will increase coverage greatly. By far the most comprehensive approach, whole genome sequencing, will accomplish this ever better, but cost is still the limiting factor for this strategy. Further, the development of large HIV/AIDS cohorts in Africa and Asia suitable for GWAS and genome sequence analyses will be welcome additions to the screening of variants for important clinically relevant ARGs [48-50].

Finally, genomics is but one of several players in AIDS research and there are some extremely promising
developments outside our field that must be mentioned. First, it is now well demonstrated that circumcision is a powerful protective aspect of HIV acquisition that is being implemented widely in the developing world to decrease the spread of HIV [66,67]. Second, three recent clinical trials have demonstrated a powerful near absolute protection against HIV infection among high-risk gay men and heterosexuals who took a single daily combination anti-retroviral pill (such pre-emptive treatment strategies are termed pre-exposure prophylaxis (PrEP)) [68,69]. In a large global cohort of 2,499 men who have sex with men, those treated with Truvada (a combination of two anti-retroviral agents, tenofovir and emtricitabine) showed a 48% reduction in HIV acquisition; this jumped to 95% protection when only those whose drug-taking compliance was 100% were considered [68]. Similar results were observed in an African trial where PrEP-treated heterosexual at-risk women were studied [69]. Anti-retroviral microbiodes are also making progress: two recent trials demonstrated 80% reduction in transmission among compliant study participants [70]. Last, in what may be his greatest legacy, former US President George W. Bush established the President’s Emergency Plan for AIDS Relief (PEPFAR) in 2002, offering US$15 billion in federal aid for treatment of HIV-infected people across the developing world. He added US$48 billion to the program in 2008. In 2002 there were less than 300,000 people on HAART in the developing world but today, thanks to PEPFAR, there are 6.6 million people on HAART. President Obama has continued PEPFAR funding and there is now tangible hope for slowing the spread of this contagion through education, treatment and research. By interdisciplinary application of important advances, the real dream is to one day see a generation of people that are free from AIDS.

Abbreviations
ARG: AIDS restriction gene; CCR5, C-C chemokine receptor; CHAVI: Center for HIV/AIDS Vaccine Immunology; GWAS: genome-wide association study; HAART: highly active anti-retroviral therapy; HIV: HIV dependency factor; LDR: linkage disequilibrium; NEMP: nuclear encoded mitochondrial protein; PEPFAR: President’s Emergency Plan for AIDS Relief; PrEP: pre-exposure prophylaxis; SNP: single nucleotide polymorphism; WNV: West Nile virus.

Author details
1Theodosius Dobzhansky Center for Genome Bioinformatics, 41A Sredny Avenue, St Petersburg State University, St Petersburg, 199034, Russia.
2Department of Biology, Shepherd University, P.O. Box 5000, Shepherdstown, West Virginia 25443-5000, USA.

Published: 31 January 2013

References
O'Brien and Hendrickson Genome Biology 2013, 14:201
http://genomebiology.com/2013/14/1/201


