Human genes that limit AIDS

Stephen J O'Brien & George W Nelson

Discernable genetic variation among people and populations has an important role in infectious disease epidemics, including that of acquired immune deficiency syndrome (AIDS). Genetic association analysis of several large AIDS cohorts implicate 14 AIDS restriction genes, polymorphic variants in loci that regulate HIV-1 cell entry, acquired and innate immunity, and cytokine defenses to HIV-1. The influence and translational impact of these genes on individual and population sensitivity to AIDS is considerable.

The scourge of AIDS, discovered three decades ago, is the most devastating epidemic of our time. The cumulative death toll is approaching 25 million, and the World Health Organization estimates that 42 million people today are infected with human immunodeficiency virus (HIV), a virus that, left untreated, kills 90% of its victims. Although considerable progress has occurred in the development of therapy for AIDS, the 16 available antiretroviral drugs are not always effective, are frequently toxic and do not clear virus from sequestered tissue reservoirs where HIV proviral genomes capable of rebound can linger for decades[1–3]. The strong drugs that slow the progress of AIDS are only now being distributed in the developing world, where they are most needed. HIV's replication efficiency, broad tissue dissemination and mutational plasticity in infected people has made vaccine development a considerable and still unsolved challenge[4–6]. For Africa, where 70% of the world's HIV-positive individuals live and die, United Nations Secretary General Kofi Annan recently named HIV/AIDS "one of the main obstacles of development itself"[7].

Although AIDS is not generally considered a genetic disease, the considerable heterogeneity in the epidemic is at least partially determined by variants in genes that moderate virus replication and immunity[5,8,9]. Not all people exposed to HIV-1 become infected, and those who do progress to AIDS-defining pathology at different time intervals. Rapid progressors succumb to AIDS in 1–5 years whereas long-term 'nonprogressors' avert AIDS for up to 20 years. There are multiple AIDS-defining conditions, including Kaposi's sarcoma, lymphoma, *Mycobacterium* tuberculosis, *Pneumocystis carinii* pneumonia (PCP) and cytomegalovirus infections. Infected individuals have heterogeneity in the strength of their innate, humoral and cell-mediated immune responses as well as differences in how they respond to antiretroviral treatment.

Stephen J. O'Brien is in the Laboratory of Genomic Diversity, National Cancer Institute, Frederick, Maryland, USA. George Nelson is in the Laboratory of Genomic Diversity, Basic Research Program, SAIC-Frederick, National Cancer Institute, Frederick, Maryland, USA.
e-mail: obrien@ncifcrf.gov

AIDS restriction genes

In the early 1980s, geneticists at the US National Cancer Institute's Laboratory of Genomic Diversity initiated a program to search the human genome for AIDS restriction genes (ARGs; human genes with polymorphic variants that influence the outcome of HIV-1 exposure or infection[10,11]). We began by enlisting collaboration with epidemiologists who were establishing prospective AIDS cohort studies. These included American people in groups at risk for exposure to HIV-1: gay men, hemophiliacs exposed to clotting factor contaminated with HIV-1 before the introduction of HIV-1 antibody blood screening in 1984 and intravenous drug users who shared needles in areas of high HIV-1 incidence (**Table 1**). Clinical, virological and immunological data were assessed twice yearly for each individual through the course of his or her infection. We transformed viably frozen lymphocytes from study participants with Epstein-Barr virus to produce lymphoblastoid cell lines, a renewable source of DNA for population-based DNA assessments. Almost 8,500 study participants comprise our present study population, 6,007 with lymphoblastoid cell lines and the balance with nonrenewable DNA specimens (**Table 1**).

We selected candidate ARGs from the growing body of host factors nominated to be involved in HIV-1 pathogenesis (**Fig. 1**). We resequenced selected human genes and adjacent regulatory regions to detect common single-nucleotide polymorphism (SNP) variants to genotype in the cohort populations, and we examined case-control comparisons of SNP allele and genotype frequencies and genetic and linkage equilibrium tests to uncover genetic influences. For ARGs that influenced the rate of AIDS onset in individuals infected with HIV-1, we used Kaplan-Meyer survival curves and Cox proportionate hazard statistics to examine the effect of different genotypes on specific AIDS outcomes[12–14].

There were many rationales for identifying ARGs, but we anticipated three principal applications. First, determining the effects of ARGs in individuals with AIDS would connect laboratory-identified host factors to AIDS pathogenesis explicitly, particularly when a plausible functional or physiological mechanism was found. Second, because the powerful antiretroviral therapy is neither universally effective nor a true cure, genetic retardation of AIDS progression could implicate new cellular targets for anti-AIDS therapy that complement the currently available antiretroviral compounds[15]. Third, ongoing clinical trials for new vac-

Table 1 Natural history of AIDS cohorts

Cohort	Principal investigators	Enrollment date	Number of individuals	Number of LCLs*	Reference
Homosexuals					
Multicenter AIDS Cohort Study	J. Phair, A. Munox, C. Rinaldo, R. Detels, R. Kaslow	1984	3,234	2,144	82–84
San Francisco City Clinic Men's Study	S. Buchbinder	1978	295	225	85
DC Gay Men's Study	J. Goedert, T. O'Brien	1982	409	296	86
Hemophiliacs					
Hemophilia Growth and Development Study	E. Gomperts, S. Donfield	1988–1989	375	320	87
Multicenter Hemophilia Cohort Study	J. Goedert, T. O'Brien	1984	2,362	1,299	88
Intravenous drug users					
ALIVE Study	D. Vlahov, S. Strathdee	1987	1,819	1,699	89
Total			8,494	6,007	

*LCLs, lymphoblastoid cell lines. Self-reported ancestry: 63% European American, 29% African American. Self-reported ethnicity: 16% Hispanic American. Hispanic Americans may have European, African or Native American ancestry, or any combination.

cines and anti-AIDS drugs, like longitudinal cohort studies, show that these drugs elicit heterogeneous responses from people with AIDS. As some portion of this heterogeneity is attributable to the distribution of ARG genotypes in the study population, a quantitative ARG score, a 'genetic propensity index' (GPI) for individuals that could be used to adjust (or subtract) the genetic 'noise' in clinical trials based on each participant's ARG genotype[16], would be useful.

Figure 1 Illustration of how HIV-1 enters cells, suggesting candidate ARGs for inspection. When HIV-1 infects a person, it seeks out tissue cell compartments where it can replicate, including lymph nodes, neural tissue, epithelium in gut or vagina, spleen, testes, kidneys and other organs. HIV's principal factories are macrophages, monocytes and T-lymphocytes, all of which carry the CD4 cell surface protein. HIV-1 enters various cells by co-opting two receptor proteins on the cell surface[92,93]. The CD4 molecule acts as a docking station, which hooks the HIV surface envelope protein and leaves the virus hanging atop the cell. Then a large seven-transmembrane-spanning cell-surface receptor, CCR5, which floats around the fluid-like cell surface membrane, meanders into contact with the CD4-snagged HIV. The CCR5-HIV interaction causes the HIV gp41 protein to change shape and penetrate the cell membrane; the HIV particle then fuses with the cell membrane and infects the cell. Infected individuals produce one to ten billion virus particles each day throughout the course of infection. The median time to CD4 cell collapse and onset of AIDS-defining pathologies is 10 years. In most infected people in Europe and USA (where HIV-1 clade B is predominant), a mutational shift occurs in *env* (called R5 to X4 tropism), which alters coreceptor preference from CCR5 to CXCR4, usually coincident with CD4 depletion. The normal chemokine ligands specific for CCR5 (RANTES, MIP1α and Mip1β) and CXCR4 (SDF1) physically block CCR5 (R5) and CXCR4 (X4) HIV-1 infection, respectively, by covering their entry coreceptors[6,23].

Since the first ARG variant, *CCR5 Δ32*, which effectively blocks HIV-1 infection in homozygous people and slows AIDS progression in heterozygotes[17–19], was reported, our group and others have documented thirteen additional ARGs through genetic association in these same cohort populations (**Table 2**). Their genetic influences differ in several ways, including mode (dominant, codominant, recessive), stage at which they act (infection, AIDS progression, the type of AIDS-defining pathogenic disease) and the interval of HIV-1 progression during which the influence is apparent. Strategies for discovering ARGs and evidence for genetic association have been reviewed[9,14], as have the physiological mechanisms of ARG action[8,16], variation in ARG frequency and influence among ethnic groups[20,21] and replication in independent cohort populations[8,14,22–24]. ARGs are polygenic and multifactorial (*i.e.*, they must interact with environmental factors, such as HIV-1 exposure, in order to be detected).

In this review, we explore the hallmarks of genetic association that indicated that these genes influence AIDS pathogenesis and the translational benefits of ARGs to therapy development. We discuss the quantitative influence of individual and composite ARG genotypes in predicting AIDS outcomes for a person with HIV and for a study population. We estimate the influence that known ARGs have on the epidemic and compute a new information theory–based epidemiological parameter, the 'explained fraction' (EF, **Box 2**) to quantify the overall influence of multiple ARGs on the rate at which a population progresses to AIDS (ref. 25 and G.W.N. & S.J.O., unpublished data). The EF offers a quantitative yardstick that estimates the proportion of epidemiological variance in AIDS progression (or in any complex multifactorial disease) contributed by one or more ARGs in the study population.
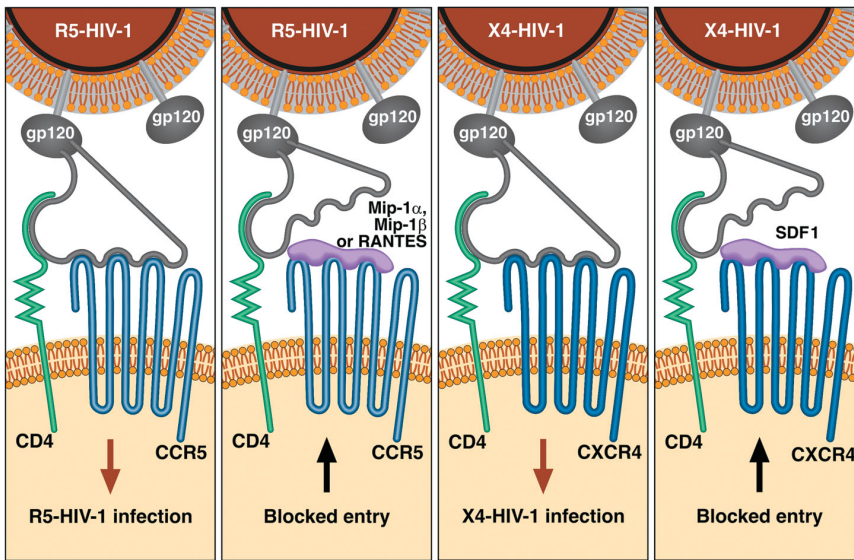
Validating ARGs in association studies

The ARGs in **Table 2** were identified and validated by genetic association studies comparing alternative disease categories (*e.g.*, infected with HIV-1 versus exposed to HIV-1 but not infected, or fast progressors versus slow progressors) for departure of allele, genotype or linkage equilibrium using the expectations of population genetic theory. Association analyses can be misleading when statistical significance alone is the basis for gene identification, and there are too many examples of artifactual associations in the literature from studies that are underpowered statistically, are inappropriately structured with respect to ancestral population admixing or contain sampling errors[26–28]. To avoid mistaken associations, several authors have proposed benchmarks or stringent criteria for validation[28,29].

For ARGs, there are five common aspects of a credible genetic association: (i) High number of clinical cases, leading to low *P* values for significance after statistical correction for multiple tests. Power calculations suggest that more than 400 cases are adequate to detect a genotypic association with a relative risk (RR) ≥2.0. (ii) Independent replication of significant associations in different populations, in different ethic groups, in different risk groups, in different cohorts and by different laboratories or investigators. If ARG allele frequencies differ among ethnic groups, the analysis should be stratified to avoid population structure artifacts[30,31]. (iii) Epidemiological influences that have high RR, high relative hazard (RH) and high attributable risk (AR, **Box 1**). Low risks mean the gene is either relatively unimportant or a statistical fluke. (iv) Gene associations assessed alone, as well as with a statistical adjustment for other previously associated gene types. This latter consideration tests epidemiological independence of multiple ARGs and uncovers epistatic interaction of genes with related physiological roles. (v) Plausible functional explanations that connect the operative SNP to a quantitative difference in a gene product and relate the change to AIDS development. Although each of these five criteria were not fulfilled for every ARG in **Table 2**, most ARGs met most of criteria. Their functional connections to AIDS progression are summarized in the next section.

Mechanism of action for ARGs

AIDS is primarily a disease of the immune system caused by a virus evolved to infect, take over and debilitate the host's highly evolved innate and acquired immune defenses. Naturally occurring genetic variants in genes empirically connected to AIDS pathogenesis were discovered in the two decades of AIDS dissemination and study. The genetic epidemiological associations have been interpreted in the context of an implicated gene's role in development of AIDS, and in each case a plausible functional mechanism has led to varying levels of empirical support (**Table 2**).

The steps in establishing a chronic HIV-1 infection leading to AIDS are illustrated in **Figure 1**. Genes encoding cellular molecules involved in the process were candidate ARGs. *CCR5* Δ32 is a naturally occurring knockout deletion (32 bp) variant. Homozygotes resistant to R5-HIV-1 infection (the principal infecting HIV-1 strain) lack the requisite HIV-1 entry coreceptor CCR5 on their lymphoid cells[8,17]. Heterozygotes express less than half the wild-type levels of CCR5 receptors, slowing HIV-1 replication, spread and pathogenesis[32,33]. The *CCR5* P1 allele, a composite haplotype comprising 13 distinct SNP alleles in the upstream promoter region of *CCR5*, confers recessive more rapid progression to AIDS[34,35]. Analysis of *CCR5* promoter alleles found no quantitative differences in luciferase transcription, HIV-1 binding or HIV-1 infectivity, although the alleles influence CCR5 abundance on lymphoid cells[36,37]. Further, oligonucleotides that define the *CCR5* P1 allele showed allele-specific recognition of nuclear transcription factors belonging to the cREL family, a result that suggests, but does not prove, that *CCR5* promoter alleles could differentially react and respond to transcription factors in different cell types[37].

The *CCR2* V64I variant mediates delay in AIDS progression[38,39] indirectly, as it causes no allele-specific quantitative differences in the amount of CCR2 produced, in the ability of alternative allelic products to bind HIV-1 or in signal transduction with CCR2-specific ligands[40,41]. One report indicated that CCR2 I64 protein can preferentially dimerize with CXCR4 polypeptides (the HIV-1 receptor that

Table 2 Genes that limit AIDS

Gene	Allele	Mode	Effect	Mechanism of action	Reference
HIV entry					
CCR5	Δ32	Recessive	Prevent infection	Knockout *CCR5* expression	17
	Δ32	Dominant	Prevent lymphoma (L)	Decrease available CCR5	90
	Δ32	Dominant	Delay AIDS	Decrease available CCR5	17
CCR5	P1	Recessive	Accelerate AIDS (E)	Increase *CCR5* expression	34
CCR2	I64	Dominant	Delay AIDS	Interact with and reduce CXCR4	38,39
CCL5	In1.1c	Dominant	Accelerate AIDS	Decrease RANTES expression	45
CXCL12	3′A	Recessive	Delay AIDS (L)	Impede CCR5-CXCR4 transition (?)	46
CXCR6	E3K	Dominant	Accelerate PCP (L)	Alter T-cell activations (?)	48
CCL2-CCL7-CCL11	H7	Dominant	Enhance infection	Stimulate immune response (?)	49
Cytokine anti-HIV					
IL10	5′A	Dominant	Limit infection	Decrease IL10 expression	53
	5′A	Dominant	Accelerate AIDS	Decrease IL10 expression	53
IFNG	−179T	Dominant	Accelerate AIDS (E)		55
Acquired immunity, cell mediated					
HLA	A,B,C	Homozygous	Accelerate AIDS	Decrease breadth of *HLA* class I epitope recognition	62,66
	B*27	Codominant	Delay AIDS	Delay HIV-1 escape	9
	B*57	Codominant	Delay AIDS	Delay HIV-1 escape	9
	B*35-Px	Codominant	Accelerate AIDS	Deflect CD8-T cell clearance of HIV-1	60
Acquired immunity, innate					
KIR3DS1	3DS1	Epistatic with *HLA-Bw4*	Delay AIDS	Clear HIV+, HLA− cells (?)	70

E, acts early in AIDS progression; L, acts late in AIDS progression; PCP, *Pneumocystis carinii* pneumonia; ?, plausible mechanism of action with no direct empirical support. All other mechanisms have quantitative expression differences among alternative alleles that suggest a mechanism.

Table 3 GPI of ARG genotypes

RH* Mode	CCR5 Δ32 Dom	CCR2 I64 Dom	CCR5 P1 Rec	CXCL12 3'A Rec	IL10 5'A Dom	CCL5 R3+R5 Dom	KIR3DS1 Dom	HLA-B*35-Px Dom	HLA-B*27 Dom	HLA-B*57 Dom	Class I	Combined GPI*
All wild-type	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
All variant	0.67	1.75	0.77	0.61	1.21	1.62	1.44	1.86	0.60	0.41	2.05	1.46
Protective	0.67	1.0	0.77	0.61	1.0	1.0	1.0	1.0	0.60	0.41	1.0	0.077
Susceptible	1.0	1.75	1.0	1.0	1.21	1.62	1.44	1.86	1.0	1.0	2.05	18.84

*RH is multiplicative across each ARG. The combined RH for an individual is defined as the GPI: RH × RH2 × RH3 × RHn for n different independent ARG genotypes, stipulating dominant, codominant or recessive genetic model, which is actually observed in the survival analysis for each ARG (**Table 2**). The GPI computation offers an empirical and deterministic approximation, providing a starting point toward predicting the survival kinetics of an individual infected with HIV-1 based on his or her composite ARG genotype. For a population of K infected people, the predicted population GPI is the geometric mean of the K composite genotypes. NA, not applicable

replaces CCR5 as an entry receptor at later stages; **Fig. 1**) whereas the wild-type CCR2 peptides do not[42]. Though it is unconfirmed, this mechanism suggests that *CCR2* V64I delays AIDS by limiting the transition from CCR5 to CXCR4 in infected individuals, a turning point in the collapse of the CD4–T lymphocyte cell population and a prelude to AIDS-defining disease[23].

RANTES is a principal chemokine ligand for CCR5. Elevated circulating RANTES levels have been detected in exposed individuals who avoid infection and also in people infected with HIV-1 who have a delayed onset of AIDS[43,44]. One of the seven SNP variants found in *CCL5* (encoding RANTES), In1.1C, is nested in an intronic regulatory sequence element that has differential allele affinity to nuclear binding proteins and whose transcription is downregulated by a factor of 4 (ref. 45). A reduction in RANTES production in individuals carrying

Figure 2 Predicted survival pattern based on GPI of four composite ARG genotypes from **Table 3**. The Weibull distribution is used to model the survival distribution, with baseline shape and scale parameters determined from AIDS cohort data, taking RH for each group from the GPI for the given genotype combination.

this allele leads to rapid AIDS progression, ostensibly by uncovering available CCR5 that facilitates the replication and spread of HIV-1.

The *CXCL12* 3'A variant is a G→A transition in the 3' untranslated region of one of two alternatively spliced transcripts of *CXCL12* (also called *SDF1*; ref. 46). Stromal-derived factor (SDF1) is the primary ligand for the late-stage HIV-1 receptor CXCR4, and the 3'A variant is 37 bp downstream of two blocks whose sequences are 88% and 92% conserved, respectively, in mouse and human transcripts[46,47]. The relatively high sequence homology is a signal for selective constraints on mutational divergence of the region because it is a putative recognition sequence for RNA- (or DNA-) binding regulatory factors. Although allele-specific cellular or virological differences have not been explicitly established, the delay in onset of AIDS observed in *CXCL12* 3'A homozygotes might result from overproduction of SDF1 in certain tissue compartments, postponing the CCR5-CXCR4 transition. A synergistic protective effect in individuals carrying *CXCL12* 3'A and *CCR2* I64 would be consistent with this model[46].

Quantitative functional allelic distinction has not been confirmed for two ARGs: *CXCR6* E3K, which is associated with late-state progression in individuals with AIDS and *Pneumocystis carinii* pneumonia, and haplotypes in the chromosome 17q gene cluster that contain three SNP variants around the chemokine genes *CCL2*, *CCL7* and *CCL11* (refs. 48,49). MCP1 (encoded by *CCL2*) and MCP3 (encoded by *CCL7*) are requisite ligands for CCR2 and CCR3, respectively, and eotaxin (encoded by *CCL11*) is a ligand for CCR3, a chemokine receptor implicated in HIV-1 infection of microglial cells in the central nervous system[50,51]. A few studies have tied MCP1, MCP3 and eotaxin to HIV-1 pathogenesis through interaction with HIV-1 directly or indirectly by stimulating cell division and migration of macrophages and dendritic cells[52]. One *CCL2-CCL7-CCL11* haplotype, H7, which includes SNP variants in a 31-kb region, is associated with resistance to HIV-1 infection, based on comparisons of highly exposed uninfected individuals with exposed infected individuals[49].

Two ARGs, interleukin-10 (*IL10*) and interferon-γ (*IFNG*), encode powerful cytokines that inhibit HIV-1 replication. The *IL10* 5'A SNP variant involves a promoter region alteration that reduces *IL10* transcription by a factor of 2–4 (ref. 53). *IL10* 5'A allele-specific synthetic oligonucleotides do not bind certain ETS-family transcription factors, which recognize the wild-type *IL10* allele sequence. Heterozygosity and homozygosity with respect to *IL10* 5'A accelerates AIDS progression, probably owing to downregulation of the inhibitory IL10 cytokine[53]. *IFNG* includes a polymorphic promoter region, one allele of which (–179T) is inducible by tumor necrosis factor (TNFα), whereas the wild-type allele (–179G) is not[54]. In African Americans, among whom the *IFNG* –179T allele has a 4% frequency (<1% in European Americans), individuals infected with HIV who are heterozygous with respect to *IFNG* –179T progress to AIDS more rapidly
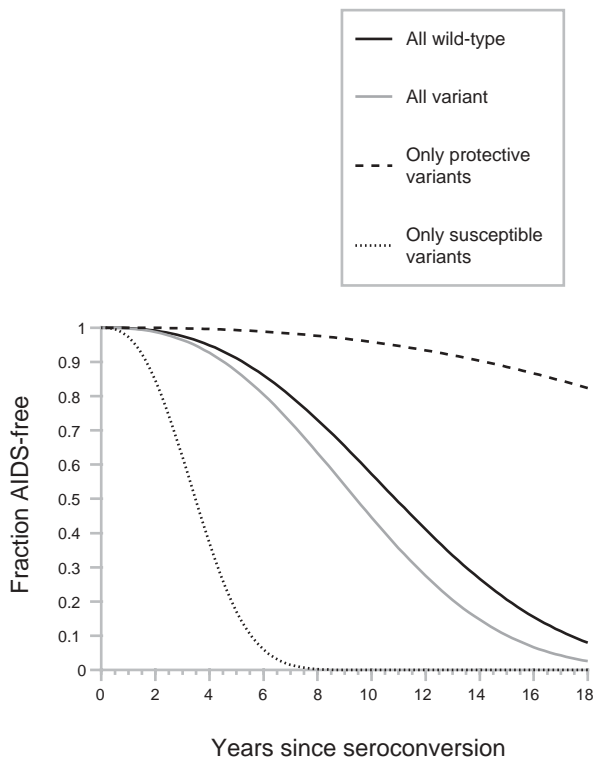
than *IFNG* –179G/G homozygotes, indicating that allele-specific inducibility confers a risk of rapid AIDS progression[55].

The essential role of *HLA*, the human major histocompatibility complex, in detecting and presenting infectious agent peptides to T-cells is well established. The *HLA* region includes 128 expressed genes, one-third of which participate in various armaments of human immunity[56]. *HLA* class I (A,B and C) and II genes (DR, DQ and DP) have considerable allele variation between individuals and populations, which is probably a reflection of the natural selective influence by historic disease outbreaks that afflicted the ancestors of modern human ethnic groups. The abundant variation in *HLA* alleles provides a broad range for individual recognition of virus agents to which they had been exposed in the past, as well as those to which they had not[57,58].

HIV-1 infects various immune cells as the prelude to HIV-1 replication, proliferation, spread and CD4-T lymphocyte damage. Because different *HLA* alleles specify cell-surface molecules with specific recognition sites for infectious agents[59], differential HIV-1 peptide motif recognition can influence both the time interval from infection to AIDS[60] and the kinetics of HIV-1 adaptive escape from immune surveillance in an infected individual[61]. For example, carriers of *HLA-B*35* experience rapid progression to AIDS and *HLA-B*35* homozygotes develop AIDS in half the median time that it takes for infected people without *HLA-B*35* to develop AIDS[62]. Molecules encoded by *HLA-B*35* comprise a serologically defined, closely related group with two distinct peptide recognition specificities. *HLA-B*35-PY* molecules recognize processed HIV peptides of nine amino acids with proline in position 2 and tyrosine in position 9; *HLA-B*35-Px* molecules present peptides with proline in position 2 and various amino acids not including tyrosine in position 9 (ref. 63). Survival association analysis indicated that *HLA-B*35*-mediated AIDS acceleration was entirely caused by *HLA-B*35-Px*, as carriers of *HLA-B*35-PY* developed AIDS at the same average rate as individuals with no *HLA-B*35* alleles[60]. The *HLA-B*35-Px* versus *HLA-B*35-PY* distinction directly connected AIDS progression to *HLA-B*35-Px* peptide recognition.

Box 1 Attributable risk of ARGs

The AR[77] quantifies the public health impact of a risk factor (in this case an ARG genotype) on the incidence of disease[78]. It is defined as the proportion of disease cases that can be attributed to an associated factor (genotype) and is related to the RR of the factor as follows:

$$AR = \frac{f\,(RR - 1)}{1 + f\,(RR - 1)}$$

where *f* is the incidence of the genotype in the study population. AR is often expressed as the fraction of disease cases that could be eliminated if the associated genotype were not present in the population. We might ask how many fewer AIDS cases would occur in the first ten years of infection if all *HLA* homozygotes were eliminated. Alternatively, what fraction of long-term survivors (people infected with HIV who avoid AIDS for ten years or longer) are in that category because they carry *CCR5* Δ32, *CCR2* I64 or another AIDS-slowing ARG genotype? **Table 4** presents individual and combined ARs for protective and susceptible ARGs.

The mechanism by which *HLA-B*35-Px* accelerates progression to AIDS does not seem to stem from a failure of the allele to stimulate cytotoxic T-cell recognition. Several amino acid residues in HIV gag proteins are *HLA-B*35-Px* targets[61], and there is an equivalent induction of cytotoxic T lymphocytes (CTLs) in carriers of *HLA-B*35-Px* as in carriers of *HLA-B*35-PY*, indicating that the requisite HIV epitopes corresponding to *HLA-B*35-Px* and *HLA-B*35-PY* stimulate CTL induction[64]. A plausible hypothesis suggests that the HIV epitopes recognized by *HLA-B*35-Px* act as an 'immunological decoy' that effectively induces CTLs that are ineffective in destroying cells infected with HIV-1, thereby

Table 4 RR, AR and EF of ARGs on AIDS progression

| | Model | Rapid progressors[a] | | | | Slow progressors[b] | | | |
		Frequency (%)	RR	AR (%)	EF (%)	Frequency (%)	RR	AR (%)	EF (%)
Protective ARGs									
CCR5 D32	Dom	19	1.10	1.8	0.1	19	1.39	6.9	0.1
CCR2 I64	Dom	18	1.15	2.6	0.0	17	1.14	2.4	0.1
CXCL12 3′A	Rec	4	1.41	1.5	0.5	4	1.54	2.0	0.7
HLA-B*27	Dom	10	1.35	3.4	1.7	10	1.70	6.3	2.4
HLA-B*57	Dom	7	1.32	2.1	0.3	6	1.51	2.9	0.4
KIR3DS1-HLA-Bw4 epistasis	Codom	–	–	–	0.9	–			1.3
Combined protective factors		54	1.36	16.2%	–	52	1.51	21.1%	–
Susceptible ARGs									
CCR5 P1	Rec	13	1.34	4.0	0.7	13	1.34	4.2	0.8
IL10	Dom	45	1.20	8.1	0.3	44	1.22	9.0	0.4
1× *HLA* Class I homozygosity	NA	17	1.15	2.4	0.2	16	1.09	1.5	0.2
2–3× *HLA* Class I homozygosity	NA	5	1.78	3.9	0.5	6	1.27	1.6	0.6
HLA-B*35-Px	Dom	9	1.66	5.9	1.4	10	1.42	4.1	1.7
CCL5 H3/H5	Dom	7	1.47	3.0	0.2	7	1.30	2.0	0.3
KIR3DS1	Codom	–	–	–	0.6	–			0.8
Combined protective factors	Dom	77	1.90	40.9%	–	78	1.37	22.4%	–
All factors EF combined	Additive	–	–	–	7.4%[c]	–	–	–	9.5%[c]
Sum of EF					7.2%				9.8%

[a]AIDS-1987 (ref. 91) < 8 y. [b]AIDS-1987 (ref. 91) >11 y. [c]We also computed the EF in the same cohorts for the influence of age on AIDS progression[80,81]. For rapid progression, EF = 2.4%; for slow progression, EF = 3.2%. The combined EF for ARGs plus environmental variable (age) = 9.9–12.7%.

monopolizing CTL defenses. In support of this possibility, carriers of *HLA-B*35-PY* reduce HIV-1 viral load far more effectively than carriers of *HLA-B*35-Px*, despite having an equivalent quantitative CTL response[64].

HLA influence on HIV-1 sequence divergence was confirmed by examining the pattern of mutation frequency among HIV-1 genomes in 473 Australians infected with HIV-1 (ref. 61). There was a dearth of sequence variants in regions of the HIV-1 genome with *HLA* recognition motifs that correspond to an infected person's *HLA* type but a mutational excess in regions with motifs recognized by *HLA* alleles that were associated with rapid progression, correlating *HLA* recognition of cognate HIV-1 motifs with AIDS survival. Two *HLA* alleles, *HLA-B27* and *HLA-B57* are consistently associated with a delayed onset of AIDS[9,65]. These *HLA* genotypes seem to be more effective in constraining HIV-1 mutational escape from immune clearance than other *HLA* alleles[9,16].

The breadth of HLA allelic and immunological diversity is good for individuals with AIDS and for populations at risk for AIDS. People who are homozygous with respect to one or more of their *HLA* class I genes (A, B or C) progress to AIDS much faster than individuals who are heterozygous with respect to *HLA-A*, *HLA-B* and *HLA-C*[62,66]. The advantage of maximal *HLA* heterozygosity in individuals (or populations) infected with HIV-1 stems from maximizing the breadth of virus epitope recognition, driven to the extreme by a swarm of HIV-1 mutational variants produced each day.

Some provocative *HLA* associations involving functionally relevant *HLA* allele groupings with HIV infection and progression have also been reported[67,68]. Certain *HLA* supertypes, grouped according to B-pocket contact residues of the antigenic peptide recognized by *HLA* class I alleles, show modest associations with HIV transmission and with circulating set-point HIV concentration (viral load)[68]. Alleles of the *HLA-Bw4* grouping, which comprises 40% of *HLA-B* alleles that share a seven–amino acid motif, are associated with long-term nonprogressors, although the study reporting this was controversial[65,67]. *HLA-Bw4* alleles serve as a ligand for certain natural killer cell receptors encoded by the killer immunoglobulin receptor (*KIR*) gene complex on chromosome 19 (ref. 69). The combination of *HLA-Bw4* and *KIR3DS1* has an epistatic protective influence on AIDS progression[70]. An interaction between the *HLA-Bw4* ligand molecules and their corresponding 'activating' receptor 3DS1 on natural killer cells seems to facilitate clearance of HIV-1-infected lymphocytes, slowing AIDS progression. In all, the immune response genes are fertile ground for discovery of important and powerful ARGs.

Implications of ARGs for AIDS therapy

The sixteen available anti-HIV-1 drugs target two enzymes encoded in the HIV-1 genome: reverse transcriptase and protease[15]. Recent drug formulations that allow single daily doses have improved compliance and prognosis, but there is still a pressing need for better, safer, cheaper and more effective anti-AIDS drugs. Defining the early steps in HIV-1

Box 2 The EF

If one considers the entire population that is at risk for a disease and all of the risk factors that mediate departure from a median survival distribution, EF quantifies the relative influence of multiple associated factors (environmental, behavioral, genetic and stochastic) and is additive until a maximum of 100% of the disease distribution is explained. EF measures our knowledge of a factor or multiple factors in predicting the occurrence of disease and EF is formulated in terms of information theory[94]. The EF is calculated directly from a two-contingency table of frequencies, where the rows represent absence or presence of a genetic factor, the columns disease cases and controls, and a_i and $a_{.j}$ are the row and column marginals. The mutual information (MI) is given by the following equation:

$$\text{MI} = I(1,2) = \sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij}\log\left(\frac{a_{ij}}{a_{i.}\,a_{.j}}\right),$$

which is simply the likelihood ratio statistic for the contingency table divided by the number of subjects *n*. *I* (1,2) (ref. 94) is the information that discriminates between dependence and independence of genotypes and disease states. EF is defined according to the following equation:

$$\text{EF} = \frac{I(1,2)}{-\sum_{j=1}^{2} D_j\log(D_j)}$$

where D_j is the frequency of cases (*i.e.*, AIDS) and controls in the study population. The denominator is the maximum mutual information, the information we would have if the genetic factors perfectly predicted the disease state. For multiple genetic factors, each combination of factors generates a row in the contingency table; the equations generalize without change to more than two rows or columns. The mathematical derivation and properties of the EF, as well as an alternative regression correlation statistic, are described in detail elsewhere (ref. 25 and G.W.N. & S.J.O., unpublished data).

EF of smoking and lung cancer. To illustrate the relationship of EF, AR and RR, consider the well-studied influence of cigarette smoking on lung cancer incidence. A 33-year follow-up study of Swedish male smokers[95] reported 36 lung cancer deaths among 8,120 nonsmokers (f = 0.0044) and 177 lung cancer deaths among 4,331 smokers (f = 0.0409). The frequency of exposure (smoking) was 0.356 and the RR is 8.93. Using equation 1, AR is 0.74; that is, 74% of lung and respiratory tract cancer deaths can be attributed to smoking. But EF from equation 3 for the same data equals 0.097, indicating that 9.7% of the occurrence or nonoccurrence of lung and respiratory tract cancers can be explained by smoking. Because EF (unlike AR) is defined for multiple exposure levels, we can assess different levels of smoking (*e.g.*, 7, 15, 25 or >25 cigarettes per day) as well as former smokers who quit. The additional information about smoking habits raises the EF for smoking to 11.5%. The reason the AR for smoking is high is that 89% of lung cancer cases were smokers, but by contrast, 98.3% of smokers never get lung cancer, indicating that factors other than smoking are also important.

---

infection and AIDS progression (**Fig. 1**) has invigorated the search for new approaches to AIDS treatment[71–74]. Specific inhibitors of HIV-1 entry into cells and integration in cellular chromosomes can block cellular factors that HIV-1 exploits in AIDS pathogenesis. After CD4, CCR5 and CXCR4 receptors were shown to catalyze HIV-1 infection (**Fig. 1**), the finding that mutational variants of receptors and their ligands stemmed HIV-1 infection and progression (**Table 2**) raised hopes that synthetic compounds that blocked the process would be suitable inhibitors of virus-cell interaction. Inhibitors of HIV-1 entry have been developed, which block CD4 binding (PRO-542 and BMS0806), CCR5 binding (SCH-C, SCH-D, PRO-140, UK-426, UK-857), CXCR4 binding (AMD-3100) and gp41-mediated membrane fusion (T-20, T-1249). All these drugs are in advanced stages of clinical trials[15], and one, T-20, was approved for public use in March 2003. In addition, an inhibitor of HIV-1 integrase (S-1360) has reached clinical setting based on efficacious results *in vitro* and in an animal model[75]. Although no human ARG that affects integration has been described to date, there is a compelling genetic evidence that the mouse locus *Fv1* resists retroviral integration[76].

New cell-based therapies have the advantage of being specific, being nontoxic, in several cases, and targeting host cell factors whose genes do not mutate and evolve resistance rapidly *in vivo* like HIV-1. Although HIV-1 can evolve resistance to cell-based therapies, the genetic restrictions of variants in *CCR5*, *CXCL12*, *CCL5* and other ARGs (**Table 2**) suggest that inhibitors of HIV-1 entry may be effective in treating infected individuals. Finally, ARG variants like *CCR5* Δ32 indicate that certain host cellular products are dispensable for normal development and immunity[8], which bodes well for cell-based therapies that target their HIV-1 involvement.

A genetic propensity index for composite ARG genotype

The interpretation of clinical trials for new vaccines and anti-HIV drugs would benefit from a method to quantify the influence a particular ARG genotype might exert on the observed heterogeneity in participants in the trial. When considering the time interval from HIV-1 infection to the onset of AIDS, it is possible to measure the influence of single or composite ARGs on AIDS survival directly in our combined AIDS cohort populations[14,16]. To do this, we computed the RH of alternative genotypes relative to the common 'wild-type' genotype (**Table 3**) based on the Cox proportionate hazards model of Kaplan-Meyer survival curves[12,13]. For multiple loci RH is multiplicative, and the combined RH estimates the GPI for an individual (**Table 3**). To illustrate the principle, we present individual RHs and composite GPIs for eleven independent (in both epidemiologic and functional context) ARG genotypes (**Table 3**) and plot the predicted AIDS survival distribution (**Fig. 2**). An arbitrary baseline of GPI = 1.0 is the composite genotype carrying the wild-type genotype at each ARG locus.

The GPI for the ARG genotype that contains all the alternative 'non-wild-type' genotypes is 1.46, not much different from baseline GPI, because the composite genotype containing all the variants includes five 'susceptible' and six 'resistant' genotypes, which offset each other in a survival distribution (**Fig. 2**). Composite GPIs for composites containing all eleven susceptible genotypes (RH ≤1.0) or all eleven resistant genotypes (RH ≥1.0) define the extremes (GPI = 0.077 and 18.84, respectively) of predicted survival times (**Table 3** and **Fig. 2**).

In a clinical trial, the GPI predicts the influence of the participants' genotypes on the trial outcome, in principle considering genotypes a confounding variable in the trial. The usefulness of such an application depends on the observed and theoretical variance of actual RH compared with GPI prediction for the study population. This variance affects sampling variance and the fraction of the overall epidemiological variance contributed by the genes we assess. Estimating this latter parameter is the next topic.

Epidemiological influence of ARGs

The AR parameter can be used to quantify the influence of any epidemiological factor on disease incidence[14,77,78]. AR considers genetic mode (dominant, recessive, etc.) disease incidence and genotype frequency in disease cases (**Box 1**). The AR for each protective ARG genotype on the rate of AIDS progression (using Equation 1) is given in **Table 4**. The AR for genotypes that slow AIDS progression ranged from 2.4% (for *CCR2* I64) to 6.9% (for
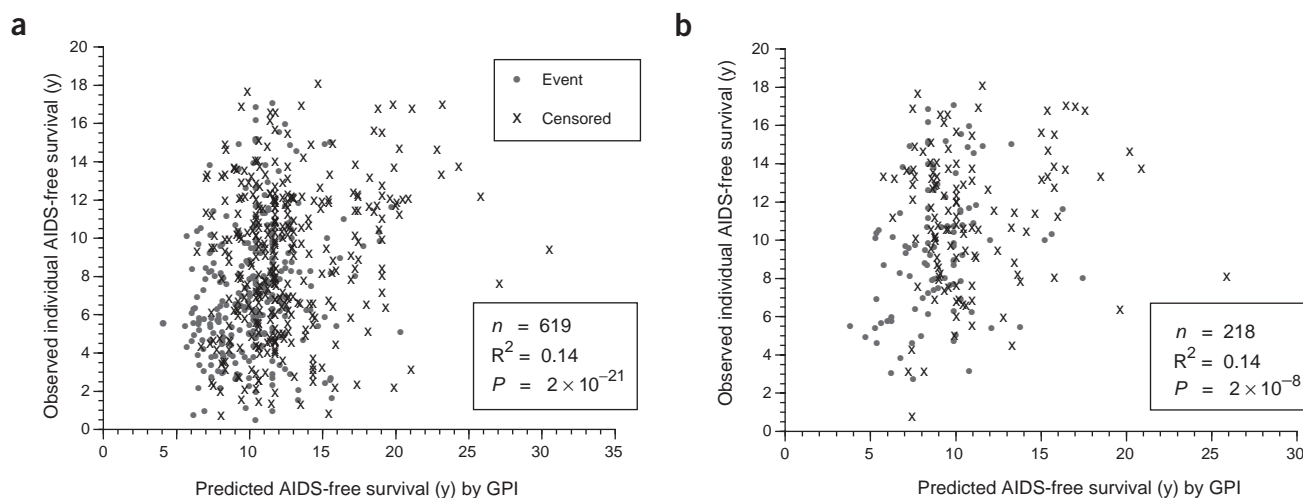

Figure 3 Ability of imputed GPI-based median survival to predict actual AIDS survival. Actual AIDS-free survival[91] versus the mean of the Weibull distribution for the subjects GPI and scale parameter, calculated by accelerated failure time regression (SAS PROC LIFEREG). Subjects are European Americans in combined AIDS cohorts (**Table 1**). (**a**) Actual survival of all subjects versus predicted survival from GPI calculated from accelerated failure time regression on all subjects. (**b**) Survival of subjects in the Multicenter Hemophilia Cohort Study and San Francisco City Clinic Men's Study cohorts versus predicted survival from GPI calculated for subjects in Multicenter AIDS Cohort Study and ALIVE Study cohorts (**Table 1**).

CCR5 Δ32) for individual genotypes. When all the genes that slow AIDS progression are considered equivalent, the cumulative AR for long-term survivors is 21.1%. This means that 21.1% of individuals infected with HIV-1 who avoided AIDS for 11 or more years did so because they carried one or more protective ARGs. Of the individuals who develop AIDS rapidly, within 5.5 years of HIV-1 infection, the single ARG effects were modest (2.4–8.1%; **Table 4**), but the cumulative AR for all the ARGs associated with rapid progression is considerable (40.9%).

Although the overall AR values seem large, there are several limitations of the AR statistic[78]. First, AR was designed for dichotomous factors, not multiple factors. This is a difficult complication because offsetting susceptible and restricting ARGs in most people effectively cancel each other's influence. Second, AR values are not inherently additive; adding AR of multiple risk factors including genotypes will often exceed 100%, distorting the actual influence. Third, AR is asymmetrical with respect to population distribution of a disease in considering only the disease cases. AR measures whether a causal factor is necessary, but does not consider whether it is sufficient, because it does not consider the incidence of the factor in nondisease categories.

For these reasons, we used an epidemiologically symmetrical parameter, the EF, to assess the overall quantitative influence of multiple ARGs or other causal factors on AIDS survival distributions (**Box 2**). EF is a mutual information-based estimate of the influence of one or more causal factors (genetic or environmental) on disease incidence in the at-risk population; EF for each factor is additive to a maximum of 100% (ref. 25 and G.W.N. & S.J.O., unpublished data). In **Table 4**, we present EFs for each individual ARG and compare the EF sum for all protective and susceptible genotype combinations with the ARs for the same genotypes. For context we also compute EFs in our cohorts (**Table 1**) for the well-known nongenetic influence of age[79,80].

The EF for each gene is small, ranging from ≤0.1% (for CCR2 I64) to 2.4% (for HLA-B*27). The EF for combined genotypes is computed for all combinations of 11 genotypes (73 different combinations). The EF for combined factors was 7.4% for the ARG influence on rapid progressors to AIDS and 9.5% for the explanation of long-term survivors who avoid AIDS for 11 years or more. The sum of the individual ARG EF estimates is 7.2% for rapid progressors and 9.8% for long-term survivors, comparable to the combined ARG genotype EF estimates (G.N. & S.J.O., unpublished data). The EF for all ARGs plus age is 9.9% for rapid progression and 12.7% for slow progression (**Table 4**).

The modest EF has important implications for the application of ARG genotype to prognosis of individuals with AIDS and to clinical trials of new therapeutics and AIDS vaccines. Because more than 90% of the epidemiological variation is not explained by the sum of ARG genotype, any weighting of study participants would have a large unexplained variance; the unexplained variance is ten times greater than the variance explained by the genes. This means that predictions based on ARGs in cohort studies or vaccine trials today would be imprecise and statistically noisy. This point is illustrated in **Figure 3**, which shows the time to AIDS predicted by GPI of 600 individuals plotted against the actual time to AIDS for the same individuals or for different cohorts. Both regressions show a modest correlation ($R^2 = 0.14$ for each; **Fig. 3**), which, though statistically significant, would hardly be useful in a clinical trial setting. As such, there is much left to explain in the varying influences on AIDS progression.

These computations do, however, offer some optimistic conclusions about AIDS and gene discovery. First, genes with modest RRs and EFs

that have measurable quantitative effects on gene expression can be validated with high statistical confidence and replicated in independent cohort designs. Second, although the identified ARGs explain only a small fraction of the variance, their EF (7.3–9.4%; **Table 4**) is comparable to that of smoking in lung cancer (9.7–11.5%; **Box 2**). Third, we applied a quantitative yardstick to measuring the additive influence of multiple genes for a complex, multifactorial disease, a useful measure that tells us how much more epidemiologic variation needs to be explained. Finally, the identified ARGs were based on SNP variation association in candidate genes, already connected to AIDS by the basic AIDS research community. As less than 20% of the ~25,000 human genes have names or suspected function, many unknown ARGs may exist.

Conclusions

The AIDS epidemic is a defining medical and public health issue of our generation. It has galvanized basic research in virology, vaccine development, antiviral therapy and health education. In this review, we attempted to highlight important points in identifying human genes that influence complex and multifactorial diseases, with the goal of applying these discoveries to further defining the regulatory components of an infected individual's cell physiology in AIDS progression and pathogenesis. The mechanism of AIDS restriction by natural genetic variants has encouraged the development of a new class of anti-AIDS drugs, which block HIV-1 binding to requisite receptors and membrane fusion (**Fig. 1**). Finally, we explore the potential application of predicting AIDS progression for a given genotype combination to inform clinical trials for new drugs and vaccines. As with genes associated with susceptibility to any complex, chronic disease, the hope is to inform diagnostics, prevention, vaccine and treatment development. For each application, the ARGs fulfill some but not all expectations, emphasizing the need for further gene identification and interpretation.

We examine two new epidemiology parameters. GPI is a quantitative score based on empirically observed epidemiological survival that predicts the RH and survival of different ARG combination genotypes (**Table 3** and **Fig. 2**). This weighting is an early step toward predicting AIDS outcome among infected individuals by considering the survival influence of the genotypes they carry. EF is an information theory–based parameter designed to estimate the portion of epidemiological variance in an epidemic that is due to a particular ARG genotype or composite multilocus genotype. A modest percentage (~9%) of the variation in AIDS progression kinetics are explained by the sum of ARGs now known, butt that fraction is comparable to the same estimate of smoking influence on lung cancer (9.7–11.5%), a well-known epidemiological factor.

The nexus of epidemiological patient cohorts and population genetic–based association approaches has shown that gene-based epidemiological analysis is a powerful tool for implicating multiple associated loci in complex disease. The identification and validation of the ARGs in **Table 2** offer hope that additional genes not yet known will be discovered not only among known candidate genes, but also by SNP haplotype-based association in high-density genome scans of clinically well-described epidemiological cohorts[81]. We estimate that more than 90% of the genetic and nongenetic influence on AIDS progression in people infected with HIV-1 is undiscovered. Nonetheless, the validation and interpretation of the 14 known ARGs lends credence to the future search by genome-wide association scans of available cohort populations. The consequences of further investigation will be valuable if their translational benefits include improved diagnostics and treatment for AIDS.

1. Palella, F.J. Jr. *et al*. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV outpatient Study Investigators. *N. Engl. J. Med.* **338**, 853–860 (1998).
2. Mocroft, A. *et al*. Changing patterns of mortality across Europe in patients infected with HIV-1. EuroSIDA Study Group. *Lancet* **352**, 1725–1730 (1998).
3. Finzi, D. *et al*. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat. Med.* **5**, 512–517 (1999).
4. Cohen, J. Shots in the Dark. The wayward search for an AIDS vaccine. (W. Norton, New York, 2001).
5. Fauci, A.S. HIV and AIDS: 20 years of science. *Nat. Med.* **9**, 839–843 (2003).
6. Weiss, R.A. HIV and AIDS: looking ahead. *Nat. Med.* **9**, 887–891 (2003).
7. Annan, K. Forward. in *AIDS in Africa* (eds. Essex, M., Mboup, S., Kanki, P.J., Marlink, R. & Tlou, S.D.) 2nd edn. (Kluwer Academic, New York, 2002).
8. O'Brien, S.J. & Moore, J. The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS. *Immunol. Rev.* **177**, 99–111 (2000).
9. Carrington, M. & O'Brien, S.J. The influence of HLA genotype on AIDS. *Ann. Rev. Med.* **54**, 535–551 (2003).
10. O'Brien, S.J. & Dean, M. In search of AIDS-resistance genes. *Sci. Am.* **277**, 44–51 (1997).
11. O'Brien, S.J. AIDS: A role for host genes. *Hosp. Pract.* **33**, 53–79 (1998).
12. Cox, D.R. & Oakes, D. *Analysis of Survival Data, London* (Chapman and Hall, 1984).
13. Allison, P.D. *Survival analyses using the SAS system: A practical guide.* 155–197 (SAS Institute, Cary, North Carolina, 1995).
14. O'Brien, S.J., Nelson, G.W., Winkler, C.A. & Smith, M.W. Polygenic and multifactorial disease gene association in man: Lessons from AIDS. *Annu. Rev. Genet.* **34**, 563–591 (2000).
15. Gulick, R.M. New antiretroviral drugs. *Clin. Microbiol. Infect.* **9**, 186–193 (2003).
16. Carrington, M., Nelson, G. & O'Brien, S. J. Considering genetic profiles in functional studies of immune responsiveness to HIV-1. *Immunol. Lett.* **79**, 131–140 (2001).
17. Dean, M. *et al*. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. *Science* **273**, 1856–1862 (1996).
18. Liu, R. *et al*. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367–377 (1996).
19. Samson, M. *et al*. Molecular cloning and functional expression of a new human CC chemokine receptor gene. *Biochemistry* **35**, 3362–3367 (1996).
20. Winkler, C.A. & O'Brien, S.J. AIDS restriction genes in human ethnic groups: An assessment. in *AIDS in Africa* (eds. Essex, M., Mboup, S., Kanki, P.J., Marlink, R. & Tlou, S.D.) 2nd edn. (Kluwer Academic, New York, 2002)
21. Winkler, C., Ping, A. & O'Brien, S.J. Patterns of ethnic diversity among the genes that influence AIDS. *Hum. Mol. Genet.* **13**, R9–R19 (2004).
22. D'Souza, M.P & Harden, V.A. Chemokines and HIV-1 second receptors - confluence of two fields generates optimism in AIDS research. *Nat. Med.* **2**, 1293–1300 (1996).
23. Berger, E.A., Murphy, P.M. & Farber, J.M. Chemokine receptors as HIV-1 coreceptors: Roles in Viral Entry, Tropism, and Disease. *Annu. Rev. Immunol.* **17**, 657–700 (1999).
24. McNicholl, J.M., Smith, D.K., Qari, S.H. & Hodge, T. Host genes and HIV: The role of the chemokine receptor gene *CCR5* and its allele (*?32 CCR5*) Emerging. *Infect. Dis.* **3**, 261–271 (1997).
25. Schemper, M. Predictive accuracy and explained variation. *Stat. Med.* **22**, 2299–2308 (2003).
26. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1517–1517 (1996).
27. Pritchard, J.K., Stephens, M., Rosenberg, N.A. & Donnelly, P. Association mapping in structured populations. *Am. J. Hum. Genet.* **67**, 170–181 (2000).
28. Glazier, A.M., Nadeau, J.H. & Aitman, T.J. Finding genes that underlie complex traits. *Science* **298**, 2345–2349 (2003).
29. Anonymous. Freely associating. *Nat. Genet.* **22**, 1–2 (1999).
30. Pritchard, J.K. & Donnelly, P. Case-control studies of association in structured or admixed populations. *Theor. Popul. Biol.* **60**, 227–237 (2001).
31. Falush, D., Stephens, D. & Pritchard, J.K. Inferences of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
32. Wu, L. *et al*. CCR5 levels and expression, pattern correlate with infectability by macrophage-tropic HIV-1, in vitro. *J. Exp. Med.* **185**, 1681–1691 (1997).
33. Benkirane, M., Jin, D.Y., Chun, R.F., Koup, R.A. & Jeang, K-T. Mechanism of transdominant inhibition of CCR5-mediated HIV-1 infection by CCR5 Delta 32. *J. Biol. Chem.* **272**, 30603–30606 (1997).
34. Martin, M.P. *et al*. Genetic acceleration of AIDS progression by a promoter variant of *CCR5*. *Science* **282**, 1907–1911 (1998).
35. Carrington, M., Dean, M., Martin, M. P. & O'Brien, S.J.: Genetics of HIV-1 infection: Chemokine receptor CCR5 polymorphism and its consequences. *Hum. Mol. Genet.* **8**, 1939–1945 (1999).
36. Salkowitz, J.R. *et al*. CCR5 promoter polymorphism determines macrophage CCR5 density and magnitude of HIV-1 propagation in vitro. *Clin. Immunol.* **108**, 234–240 (2003).
37. Bream, J.H. *et al*. *CCR5* promoter alleles distinguished by specific DNA binding factors. *Science* **284**, 223a (1999).
38. Smith, M.W. *et al*. Contrasting genetic influence of CCR2 and CCR5 receptor gene variants on HIV-1 infection and disease progression. *Science* **277**, 959–965 (1997).
39. Anzala, A.O. *et al*. CCR2-64I allele and genotype association with delayed AIDS progression in African women. *Lancet* **351**, 1632–1633 (1998).
40. Lee, B. *et al*. Influence of the *CCR2-V64I* polymorphism on human immunodeficiency virus type 1 coreceptor activity and on chemokine receptor function of CCR2b, CCR3, CCR5, and CXCR4. *J. Virol.* **72**, 7450–7458 (1998).
41. Mariani, R. *et al*. *CCR2*-64I polymorphism is not associated with altered CCR5 expression or coreceptor function. *J. Virol.* **73**, 2450–2459 (1999).
42. Mellado M., Rodriguez-Frade, J.M., Vila-Coro A.J., de Ana A.M. & Martinez-A.C. Chemokine control of HIV-1 infection. *Nature* **400**, 723–724 (1999).
43. Zagury, D. *et al*. C-C chemokines, pivotal in protection against HIV type 1 infection. *Proc. Natl. Acad. Sci. USA* **95**, 3857–3861 (1998).
44. Gallo, R.C., Garzino-Demo, A. & DeVico, A.L. HIV infection and pathogenesis: what about chemokines? *J. Clin. Immunol.* **19**, 293–299 (1999).
45. An, P. *et al*. Modulation of HIV infection and disease progression by interacting *RANTES* variants. *Proc. Natl. Acad. Sci USA* **99**, 10002–10007 (2002).
46. Winkler, C. *et al*. Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. *Science* **279**, 389–393 (1998).
47. Bleul, C.C. *et al*. The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature* **382**, 829–833 (1996).
48. Duggal, P. *et al*. Genetic influence of CXCR6 chemokine receptor alleles on PCP-mediated AIDS progression among African-Americans. *Genes Immun.* **4**, 245–250 (2003).
49. Modi, W.S. *et al*. Directional selection in the MCP-1 MCP-3 Eotaxin gene cluster in influences HIV-1 transmission. *AIDS* **17**, 2357–2365 (2003).
50. Bagiolini, M. *et al*. IL-8 and related chemotactic cytokines-CXC and CC chemokines. *Adv. Immunol.* **55**, 97–179 (1994).
51. He, J.L. *et al*. CCR3 and CCR5 are co-receptors for HIV-1 infection of microglia. *Nature* **385**, 645–649 (1997).
52. Sozzani, S. *et al*. Differential regulation of chemokine receptors during dendritic cell maturation: a model for their trafficking properties. *J. Immunol.* **161**, 1083–1086 (1998).
53. Shin, H.D. *et al*. Genetic restriction of HIV-1 infection and AIDS progression by promoter alleles of interleukin 10. *Proc. Natl. Acad. Sci. USA.* **97**, 14467–14472 (2000).
54. Bream, J.H., An, P., Zhang. X., Winkler, C.A. & Young, H.A. A single nucleotide polymorphism in the proximal IFN-gamma promoter alters control of gene transcription. *Genes Immun.* **3**, 165–169 (2002).
55. An, P. *et al*. A TNF-inducible promoter variant of interferon-gamma accelerates CD4 T cell depletion in HIV-1 infected individuals. *J. Infect. Dis.* **188**, 228–231 (2003).
56. Aguado, B. *et al*. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* **401**, 921–923 (1999).
57. Parham, P. & Ohta, T. Population biology of antigen presentation by MHC class I molecules. *Science* **272**, 67–74 (1996).
58. Hughes, A.L. & Yeager, M. Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genet.* **32**, 415–435 (1998).
59. Barber, L.D. *et al*. Overlap in the repertoires of peptides bound in-vivo by a group of related Class-I HLA-B allotypes. *Curr. Biol.* **5**, 179–190 (1995).
60. Gao, X. *et al*. Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N. Engl. J. Med.* **344**, 1668–1675 (2001).
61. Moore, C.B. *et al*. Evidence of HIV-1 adaptation to HLA-restricted immune response at a population level. *Science* **296**, 1439–1443 (2003).
62. Carrington, M. *et al*. HLA and HIV-1: Heterozygote advantage and *B*35-Cw*04* disadvantage. *Science* **283**, 1748–1752 (1999).
63. Steinle, A. *et al*. Motif of HLA-B*3503 peptide ligands. *Immunogenetics* **43**, 105–107 (1996).
64. Jin, X. *et al*. Human immunodeficiency virus Type 1 (HIV-1)-specific CD8+-T-Cell responses for groups of HIV-1-infected individuals with different HLA-B*35 genotypes. *J. Virol.* **76**, 12603–12610 (2002).
65. O'Brien, S.J., Gao, X. & Carrington, M. HLA and AIDS: A cautionary tale. *Trends Mol. Med.* **7**, 379–381 (2002).
66. Tang, J.M. *et al*. HLA class I homozygosity accelerates disease progression in human immunodeficiency virus type I infection. *AIDS Res. Hum. Retroviruses* **15**, 317–324 (1999).
67. Flores-Villanueva, P.O. *et al*. Control of HIV-1 viremia and protection from AIDS are associated with HLA-Bw4 homozygosity. *Proc. Natl. Acad. Sci. USA* **98**, 5140–5145 (2001).
68. Trachtenberg, E. *et al*. Advantage of rare HLA supertype in HIV disease progression. *Nat. Med.* **9**, 928–935 (2003).
69. Lanier, L.L. NK cell receptors. *Annu. Rev. Immunol.* **16**, 359–393 (1998).
70. Martin, M.P. *et al*. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat. Genet.* **31**, 429–434 (2002).
71. Cairns, J.S. & D'Souza, M.P. Chemokines and HIV-1 second receptors: The therapeutic connection. *Nat. Med.* **4**, 563–568 (1998).
72. O'Brien, S.J. A new approach to therapy. *HIV Newsline* **4**, 3–6 (1998).
73. LaBranche, C.C. *et al*. HIV fusion and its inhibition. *Antiviral Res.* **50**, 95–115 (2001).
74. Stevenson, M. New targets for inhibitors of HIV-1 replication. *Nat. Rev. Mol. Cell Biol.* **1**, 40–49 (2000).
75. Hazuda, D.J. *et al*. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **287**, 646–650 (2000).
76. Best, S., Le Tissier, P., Towers, G. & Stoye, J.P. Positional cloning of the mouse retrovirus restriction gene *Fv-1*. *Nature* **382**, 826–829 (1996).

77. Levin, M.L. The occurrence of lung cancer in man. *Acta Unio Int. Contra Cancrum* **9**, 531–541 (1953).
78. Benichou, J. Attributable Risk. in *Encyclopedia of Statistics.* (eds. Armitage, P. & Colton, T.) (Wiley, New York, 1998).
79. Goedert, J.J. *et al.* A prospective study of human immunodeficiency virus type 1 infection and the development of AIDS in subjects with hemophilia. *N. Engl. J. Med.* **321**, 1141–1148 (1989).
80. Darby, S.C., Doll, R., Thakrar, B., Rizza, C.R. & Cox, D.R. Time from infection with HIV to onset of AIDS in patients with haemophilia in the UK. *Stat. Med.* **6**, 681–689 (1990).
81. The International Hap Map Project. *Nature* **426**, 789–796 (2003).
82. Phair, J. *et al.* Acquired-Immune-Deficiency-Syndrome occurring within 5 years of infection with Human-Immunodeficiency-Virus Type-1 - the Multicenter AIDS Cohort Study. *J. Acquir. Immune Defic. Syndr.* **5**, 490–496 (1992).
83. Detels, R. *et al.* For the Multicenter AIDS Cohort Study: Resistance to HIV-1 infection. *J. Acquir. Immune Defic. Syndr.* **7**, 1263–1269 (1994).
84. Kaslow, R.A. *et al.* The Multicenter AIDS Cohort Study - Rationale, organization, and selected characteristics of the participants. *Am. J. Epidemiol.* **26**, 310–318 (1987).
85. Buchbinder, S.P., Katz, M.H., Hessol, N.A., O'Malley, P.M. & Holmberg, S.D. Long-term HIV-1 infection without immunologic progression. *AIDS* **8**, 1123–1128 (1994).
86. Goedert, J.J. *et al.* Decreased helper lymphocytes-T in homosexual men. 1. Sexual contact in high-incidence areas for the Acquired Immunodeficiency Syndrome. *Am. J. Epidemiol.* **121**, 629–636 (1985).
87. Hilgartner, M.W. *et al.* Hemophilia growth and development study - design, methods, and entry data. *Am. J. Ped. Hematol. Oncol.* **15**, 208–218 (1993).
88. Lederman, M.M. *et al.* Human immunodeficiency virus (HIV) type 1 infection status and in vitro susceptibility to HIV infection among high-risk HIV-1-seronegative hemophiliacs. *J. Infect. Dis.* **172**, 228–231 (1995).
89. Vlahov, D. *et al.* Association of drug injection patterns with antibody to human immunodeficiency virus type 1 among intravenous durg users in Baltimore, Maryland. *Am. J. Epidemiol.* **132**, 847–856 (1990).
90. Dean, M. *et al.* Reduced risk of AIDS lymphoma in individuals heterozygous for the CCR5-?32 mutation. *Cancer Res.* **59**, 3561–3564 (1999).
91. US Centers for Disease Control. *Morb. Mortal. Wkly. Rep.* **36** Suppl 1 (1987).
92. Goldsmith, M.A. & Doms, R.W. HIV entry: are all receptors created equal? *Nat. Immunol.* **3**, 709–710 (2002).
93. Littman, D.R. Chemokine receptors: keys to AIDS pathogenesis? *Cell* **93**, 677–680 (1998).
94. Kullback, S. *Information Theory and Statistics* (Dover, Mineola, New York, 1968).
95. Nilsson, S., Carstensen, J.M. & Pershagen, G. Mortality among male and female smokers in Sweden: a 33 year follow up. *J. Epidemiol. Community Health* **55**, 825–830 (2001).