

The Influence of *CCL3L1* Gene-Containing Segmental Duplications on HIV-1/AIDS Susceptibility

Enrique Gonzalez,^{1*} Hemant Kulkarni,^{1*} Hector Bolivar,^{1*} Andrea Mangano,^{2*} Racquel Sanchez,^{1†} Gabriel Catano,^{1†} Robert J. Nibbs,^{3†} Barry I. Freedman,^{4†} Marlon P. Quinones,^{1†} Michael J. Bamshad,⁵ Krishna K. Murthy,⁶ Brad H. Rovin,⁷ William Bradley,^{8,9} Robert A. Clark,¹ Stephanie A. Anderson,^{8,9} Robert J. O'Connell,^{9,10} Brian K. Agan,^{9,10} Seema S. Ahuja,¹ Rosa Bologna,¹¹ Luisa Sen,² Matthew J. Dolan,^{9,10,12‡} Sunil K. Ahuja^{1‡}

¹Veterans Administration Research Center for AIDS and HIV-1 Infection, South Texas Veterans Health Care System, and Department of Medicine, University of Texas Health Science Center, San Antonio, TX 78229, USA. ²Laboratorio de Biología Celular y Retrovirus, and ¹¹Servicio de Infectología, Hospital de Pediatría "J.P. Garrahan," Buenos Aires, Argentina. ³Cancer Research UK Beatson Laboratories, Glasgow, Scotland, G61 1BD ⁴Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, 27157, USA. ⁵Departments of Human Genetics and Pediatrics, University of Utah, Salt Lake City, UT 84112, USA. ⁶Southwest Foundation for Biomedical Research, San Antonio, TX, 78227, USA. ⁷Division of Nephrology, Ohio State University, Columbus, OH, 43210, USA. ⁸Henry M. Jackson Foundation, ⁹Tri-Service AIDS Clinical Consortium, and ¹⁰Infectious Diseases Service, Wilford Hall Medical Center, Lackland AFB, TX 78236, USA. ¹²Defense Institute for Medical Operations, Brooks City-Base, TX 78235, USA.

*These authors contributed equally to this work.

†These authors contributed equally to this work.

‡To whom correspondence should be addressed. E-mail: ahujas@uthscsa.edu (S.K.A.) and matthew.dolan@lackland.af.mil (M.J.D.)

Segmental duplications in the human genome are selectively enriched for genes involved in immunity, although the phenotypic consequences for host defense are unknown. We show that there are significant interindividual and interpopulation differences in the copy number of a segmental duplication encompassing the gene encoding *CCL3L1* (*MIP-1 α P*), a potent HIV-1-suppressive chemokine and ligand for the HIV coreceptor *CCR5*. Possession of a *CCL3L1* copy number lower than the population average is associated with markedly enhanced HIV/AIDS susceptibility. This susceptibility is even greater in individuals who also possess disease-accelerating *CCR5* genotypes. This relationship between *CCL3L1* dose and altered HIV/AIDS susceptibility points to a central role for *CCL3L1* in HIV/AIDS pathogenesis, and indicates that differences in the dose of immune response genes may constitute a genetic basis for variable responses to infectious diseases.

Duplicated host defense genes that are known to have dosage effects are thought to contribute to the genetic basis of some complex diseases, although direct evidence for this is lacking. We surmised that a hot-spot for segmental duplications on human chromosome 17q might be relevant to immunity against infectious diseases such as HIV-1, since it encompasses two CC chemokine genes, CC chemokine ligand 3-like 1 (*CCL3L1*; other names, *MIP-1 α P* and *LD78 β*)

and *CCL4L1* (*MIP-1 β -like*), which represent the duplicated isoforms of *CCL3* and *CCL4*, respectively (1–3). As a consequence of these duplications, the copy number of *CCL3L1* and *CCL4L1* varies among individuals (2, 3) (fig. S1A). This is important because *CCL3L1* is the most potent known ligand for CC chemokine receptor 5 (*CCR5*), the major coreceptor for HIV, and it is a dominant HIV-suppressive chemokine (3).

In light of this relationship between *CCL3L1* and its in vitro effect on HIV infection, we selected HIV infection as a model system in which to test our hypothesis that segmental duplications causing dosage effects of host defense genes are associated with phenotypic effects in vivo. To test this hypothesis, we determined the distribution of chemokine gene-containing segmental duplications in 1,064 humans from 57 populations and 83 chimpanzees (4). We next analyzed 4,308 HIV-1-positive (*HIV*⁺) and -negative (*HIV*⁻) individuals from groups with different geographical ancestries (e.g., Africans, Europeans) to determine if the risk of acquiring HIV, and second, the rate at which HIV disease progressed were sensitive to differences in the dose of *CCL3L1* gene-containing segmental duplications [(4); Supplementary online material (SOM) section 4.1].

Non-random distribution of *CCL3L1*-containing segmental duplications. African populations possessed a significantly greater number of *CCL3L1* gene copies than

non-Africans (Fig. 1 and fig. S1B). The geographic region-of-origin explained nearly 35% of the total variation in the distribution of *CCL3L1* gene copies (analysis of variance: $F = 94.41$, $df = 6$, 1037 ; $P = 1.23 \times 10^{-94}$). Corroborating this, in separate cohorts of HIV⁻ subjects, there were significant inter-individual and inter-population differences in *CCL3L1* copy numbers. The median copy number in HIV⁻ Argentinean children was two, and in HIV⁻ African (AA)-, European (EA)-, and Hispanic (HA)-American adults, it was four, two, and three, respectively (Fig. 2, A to D, open bars, and fig. S2).

The duplicated region encoding human *CCL3L1* had an ancestral correlate in chimpanzee (Fig. 1 and fig. S3). Together, these results demonstrated that there were significant differences between species and among human populations in the frequency of chemokine gene-containing segmental duplications (Fig. 1, B and C). Nevertheless, despite these differences, the dispersion around the average copy number was similar in both human populations and chimpanzees (Fig. 1B and fig. S1B). Based on these observations, we hypothesized that it is not the absolute copy number per se, but rather the gene dose relative to the average copy number in each population that confers HIV/AIDS susceptibility.

***CCL3L1* gene dose and HIV/AIDS susceptibility.**

Several lines of evidence, from four different human populations and in the setting of two different modes of acquiring HIV, i.e., mother-to-child and adult-to-adult, indicated that possession of a low *CCL3L1* copy number was a major determinant of enhanced HIV susceptibility among individuals. Individuals with a low *CCL3L1* copy number were overrepresented among the HIV⁺ compared with HIV⁻ subjects (shift to the left in Fig. 2, A to D, and figs. S2 and S4). Based on the consistency, strength and significance of the differences in the distribution of *CCL3L1* copy numbers in the HIV⁺ and HIV⁻ individuals in each of the cohorts studied, we rejected the null hypothesis of no association between risk of acquiring HIV and *CCL3L1* copy number (Fig. 2, A to D and fig. S2).

We next determined the strength of the association between *CCL3L1* copy number and risk of acquiring HIV (Fig. 2, E to H). In our initial analyses, we chose the population-specific median copy number in the uninfected group as a reference point to compute the risk of acquiring HIV (SOM section 5.1). Compared with possession of two copies of *CCL3L1*, children possessing <2 or >2 copies had significantly higher or lower risks, respectively, of acquiring HIV (Fig. 2E). This association was evident in the analysis of the entire cohort of children with (table S1) or without (Fig. 2E) adjustments for receipt of zidovudine prophylaxis given to reduce the risk of transmission and for individuals who received no prophylaxis (table S1). Notably, with each

increase in *CCL3L1* copy number above the median, there was a dose-dependent, step-wise decrease in the risk for acquiring HIV (Fig. 2E). The findings depicted in Fig. 2 (F to H), and those derived from a separate analysis in another cohort of 1,133 HIV⁻ individuals matched for ethnicity/race (fig. S2), indicated that adults who possessed a *CCL3L1* copy number lower than the population-specific median were at a higher risk of acquiring HIV. Thus, in each population, the median number of *CCL3L1* copies served as the transition point at which the balance tilted in favor of protection against acquiring HIV.

We also estimated the risk of acquiring HIV across the cline of population-specific high to low *CCL3L1* copy numbers (fig. S4). Depending on the study population, each *CCL3L1* copy lowered the risk of acquiring HIV by 4.5–10.5%, indicating that the population-specific high and low *CCL3L1* copy numbers are at different ends of a distribution of HIV susceptibility (SOM section 5.2). Substantiating this, relative to possession of the population-specific high *CCL3L1* copy numbers shown in fig. S4, individuals who had a low copy number had between 69 and 97% higher risk of acquiring HIV (fig. S4).

The aforementioned analyses were conducted using logistic regression. While membership in either the HIV⁺ or HIV⁻ group is not a random outcome, to the extent that these two groups can be thought of as random samples from their respective subsets of a well-defined population, logistic regression on group membership allows estimation of the relative odds of being HIV⁺ or HIV⁻ for two different copy numbers. In this sense, logistic regression permits the determination of the association between *CCL3L1* copy number and risk of acquiring HIV. As the number of *CCL3L1* copies followed a Poisson distribution (data not shown), we also used Poisson regression analysis to determine the association between *CCL3L1* copy number and risk of acquiring HIV infection. These analyses assume that copy numbers are conditional on HIV status, and show that HIV⁺ subjects have significantly lower *CCL3L1* copy numbers than HIV⁻ subjects (Fig. 2, I). Although the true risk of HIV-acquisition associated with possession of different *CCL3L1* gene copy numbers can theoretically only be estimated from a longitudinal study, in the cohorts we studied, the results of two different statistical approaches demonstrate a strong association between possession of low *CCL3L1* copy number and risk of acquiring HIV infection (Fig. 2, E and J, and table S1).

In addition to influencing HIV acquisition, the number of *CCL3L1* copies was associated with variable rates of disease progression (figs. S5 and S6). For example, in the adult HIV⁺ cohort, a gene dose lower than the overall cohort median or population-specific median was associated with a dose-dependent increased risk of progressing rapidly to AIDS or

death (Fig. 3, A and B, and figs. S5). A disease-influencing effect of *CCL3L1* dose was not detected in the HIV⁺ children, suggesting that either the roles of *CCL3L1* in HIV⁺ adults and children differ, or that the short follow-up time in the pediatric cohort was insufficient to detect an effect.

Mechanistic links between *CCL3L1* dose and HIV/AIDS susceptibility. Increasing *CCL3L1* copy number was positively associated with CCL3/CCL3L1 secretion and negatively associated with the proportion of CD4⁺ T cells that express CCR5 (Fig. 3, C and D) (2). Additionally, there was a dose-dependent association between *CCL3L1* copy number and the viral set point and rate of change in CD4⁺ T cell counts, two well-established predictors of clinical outcome (5); low *CCL3L1* doses were associated with a higher viral set point and greater subsequent T cell loss (Fig. 3, E and F). These relationships might explain the association between *CCL3L1* gene dose and risk of acquiring HIV and disease progression as (i) chemokines are thought to mediate their HIV-suppressive activity by steric blocking of the interaction between gp120 and CCR5 or ligand-mediated internalization of CCR5, reducing its availability for use by gp 120 ((3) and references therein), and (ii) high CCR5 ligand and/or low CCR5 receptor expression represents a correlate of HIV/AIDS protection ((6-12) and references therein).

Phenotypic equivalency of population-specific *CCL3L1* gene doses. Human populations differ in their *CCL3L1* gene content (Fig. 1). Accordingly, it was important to determine whether an absolute *CCL3L1* copy number (e.g., 2 copies) was associated with similar transmission- and/or disease-influencing phenotypic effects in different populations. To do so, we compared the associated phenotypic effects of similar and dissimilar *CCL3L1* copy numbers in HIV⁺ EAs and AAs (Fig. 3, G to N), and the change in the frequency distribution of copy number in these two populations over time (Fig. 3, O and P). The findings indicated that in HIV⁺ EAs and AAs, the *CCL3L1* copy number that corresponded to the population-specific median, half-median and low doses (i) were associated with comparable rates of disease progression or changes in CD3⁺, CD4⁺ or CD8⁺ T cell counts (Fig. 3, G to N, and table S2), and (ii) had similar trajectories with respect to the changes in their distribution profiles over time (Fig. 3, O and P and figs. S7 and S8). By contrast, possession of two *CCL3L1* copies, i.e., the median and half-median gene dose in EAs and AAs, respectively, was associated with differing rates of disease progression (Fig. 3K). Consistent with this finding, the trajectories of the change in the frequency distribution of individuals possessing two *CCL3L1* copies differed over time: increasing in HIV⁺ EAs, but declining in HIV⁺ AAs (Fig. 3, O and P). These findings, together with those shown in Fig. 2 and SOM section 5.1, collectively support the concept that different *CCL3L1* gene doses among populations are associated with phenotypically equivalent

effects (Fig. 3Q). They also imply that the phenotypic effects associated with *CCL3L1* gene dosage cannot be estimated by knowing only the absolute *CCL3L1* copy number. This value, in any given individual, is meaningful only if compared to the distribution of *CCL3L1* copies in the geographic ancestral population of the given individual (SOM section 5.1).

Distribution of *CCL3L1* gene copies under HIV selective pressure. The association between *CCL3L1* gene dose and HIV/AIDS susceptibility in adults (Figs. 2 and 3, A and B) predicts that the following pattern should be discernable in a prospective longitudinal cohort in which subjects are recruited at an early stage of infection. Initially, the HIV⁺ cohort will be enriched for individuals with *CCL3L1* copy numbers lower than the population-specific median. Over time, the prevalence of these individuals will decrease because of their rapid progression to AIDS/death. As a result, the prevalence of HIV⁺ subjects with *CCL3L1* copy numbers equal to or greater than the population-specific median will increase. Thus, with increasing follow-up times, the distribution of *CCL3L1* copies will begin to resemble that found in HIV⁻ subjects. The value of testing this prediction is that it combines, into a single analytical model, analyses of (i) the susceptibility to infection in individuals with different numbers of *CCL3L1* copies, and (ii) the time-to-equilibrium between the virus and *CCL3L1* genotype-dependent events in the infected host. Our results are consistent with these predictions (Fig. 3, O and P, and figs. S7 to S9). These observations suggest that infection with HIV-1 can exert a negative selective pressure on individuals with low copy numbers which, depending on the strength of this effect in the general population, could change the population-specific distribution of *CCL3L1* copy number.

***CCL3L1* dose and CCR5 genotypes in HIV/AIDS susceptibility.** We and others have shown that *CCR5* haplotypes that include *CCR5* promoter polymorphisms as well as coding polymorphisms in *CCR2* (*CCR2-V64I*) and *CCR5* (*A32*) influence the risk of acquiring HIV and disease progression ((12-15) and references therein). However, *CCR5* is part of a complex system in which virus interacts with *CCR5* and *CCR5* interacts with various ligands. Thus, if gene-gene interactions are not considered, these interactions might complicate analysis of the in vivo contributions of *CCR5* genotypes. This concern is made all the more apparent by the observation that *CCR5* protein expression levels are influenced not only by variants in *CCR5* ((16, 17) and references therein), but also by *CCL3L1* (Fig. 3C). Thus, virus X *CCR5* X *CCL3L1* interactions in vivo and the phenotypic effects associated with *CCR5* genotypes could depend, in part, on the genetic background conferred by *CCL3L1* dose. To test this hypothesis, we determined the phenotypic effects attributable to *CCL3L1* gene dose alone,

CCR5 haplotype pairs (genotypes) alone, and their combined effects.

The HIV⁺ adult cohort was stratified into four mutually exclusive genetic risk groups (GRGs) based on possession of population-specific, low or high number of *CCL3L1* copies (*CCL3L1*^{low} or *CCL3L1*^{high}) and disease-accelerating, i.e., detrimental (det) or non-detrimental (non-det) *CCR5* genotypes (*CCR5*^{det} or *CCR5*^{non-det}; Fig. 4A). Of the four GRGs, *CCL3L1*^{high}*CCR5*^{non-det} and *CCL3L1*^{low}*CCR5*^{det} were at the two extremes of HIV/AIDS susceptibility (Fig. 4, B to I). Relative to possession of *CCL3L1*^{high}*CCR5*^{non-det}, *CCL3L1*^{low}*CCR5*^{det} was associated with a ≥ 3-fold greater risk of progressing rapidly to 8 of 12 AIDS-defining illnesses (Table 1). By contrast, the *CCL3L1*^{high}*CCR5*^{det} and *CCL3L1*^{low}*CCR5*^{non-det} genotypes were associated with a ≤ 3-fold higher risk of progressing to three or four of these 12 illnesses, respectively (Table 1).

The trajectory of the frequency distribution profiles of the four *CCL3L1/CCR5* GRGs in individuals with varying follow-up times were also revealing in that they closely paralleled those described previously for a variable number of *CCL3L1* copies alone (compare Fig. 4J vs Fig. 3, O and P, and figs. S7 to S9). Thus, significant changes occurred only in the frequencies of the two GRGs that contained *CCL3L1*^{low} and *CCL3L1*^{high}*CCR5*^{non-det}, such that over time the distribution of the GRGs in surviving HIV⁺ subjects approached ever closer to the values observed in the HIV⁻ population (Fig. 4, J to L).

Taken together, in the context of a well-characterized prospective cohort comprising of HIV⁺ EAs and AAs, the *CCL3L1/CCR5*-based genomic signature for HIV/AIDS susceptibility was *CCL3L1*^{low}*CCR5*^{det} > *CCL3L1*^{low}*CCR5*^{non-det} ≥ *CCL3L1*^{high}*CCR5*^{det} > *CCL3L1*^{high}*CCR5*^{non-det}. This observation implied that *CCL3L1*^{low} may have a stronger effect than disease-accelerating, detrimental *CCR5* genotypes in influencing HIV/AIDS pathogenesis in these two populations. Additionally, these findings suggest that a population-specific low *CCL3L1* dose provides a permissive genetic background for the full expression of the phenotypic effects associated with detrimental *CCR5* genotypes. This was apparent because (i) relative to genotypes that contained only *CCR5*^{det}, those that contained *CCL3L1*^{low} with or without *CCR5*^{det} were associated with a higher risk of acquiring HIV (compare green vs orange or red color-coded GRGs in Fig. 4, H and I); and (ii) the maximal disease-accelerating effects associated with detrimental *CCR5* genotypes occurred mainly in individuals who also possessed a low number of *CCL3L1* copies relative to the population-specific average (compare Kaplan–Meier plots for *CCL3L1*^{high}*CCR5*^{det} and *CCL3L1*^{low}*CCR5*^{det} in Fig. 4, E and F).

Public health impact of variations in *CCL3L1* and *CCR5*. In the populations examined, up to 42% of the burden

of infection and ~30% of the accelerated rate of progression to AIDS were attributable to variations in *CCL3L1/CCR5* (black bars in Fig. 5 and fig. S11). The largest contributor to the burden of HIV/AIDS was possession of a population-specific low *CCL3L1* copy number (Fig. 5, compare combination of red and orange to green bars, and fig. S11). These findings suggest that the contribution of *CCL3L1* copy number is comparable to, or more than, *CCR5* genotype in influencing the epidemiology of HIV in the populations examined. These results also substantiate the observation that the disease-accelerating effects associated with variation in *CCR5* depend, in part, on the genetic background of *CCL3L1* copy number.

Discussion. These findings have five major implications. First, they provide a precedent for a link between segmental duplication events leading to changes in the dose of an immune response gene and variability in the phenotypic response to an infectious disease. Recent human–non-human primate comparative genomic analyses have led to the prediction that genes embedded within segmental duplications might have enhanced the ability of humans to adapt to their environments (1, 18). Our findings support this prediction.

Second, *CCL3L1* gene dose is a novel means of buffering against the risk of HIV infection and/or disease progression in the populations examined. *CCL3L1* gene doses lower than the population-specific average provides a genetically “unbuffered” state with respect to the risk of HIV/AIDS susceptibility. However, it is important to emphasize that it is not the absolute gene copy number per se, but the copy number within the overall genetic context that confers phenotypic expression. This genetic context varies among populations as a result of their different demographic and evolutionary histories. Thus, an individual’s specific *CCL3L1* gene dose and *CCR5* genotype are associated with susceptibility to HIV/AIDS, but only when viewed in the context of that person’s geographical ancestry (Fig. 3Q) (14).

Third, within the populations examined, the Bradford–Hill criteria (19) for causality between *CCL3L1* dose and risk of acquiring HIV were met (SOM section 5.3). Thus, by analogy to the genetic studies that established the paradigm of “no *CCR5*–no HIV-1 infection,” the current findings establish that of “*CCL3L1*^{low}–enhanced HIV/AIDS susceptibility”. These findings provide strong genetic underpinnings for the substantial body of evidence that *CCR5* ligands play an important anti-HIV-1 role in vivo ((20) and references therein). Paradoxically, they also indicate that a network of HIV-suppressive *CCR5* ligands (e.g., RANTES) cannot fully compensate for the functional state conferred by *CCL3L1*^{low}. Therefore, *CCL3L1*-mediated immune responses may be required to thwart HIV infection and the complications that occur during HIV-induced immune suppression.

Fourth, *CCL3L1* gene dose may be an important genetic correlate of vaccine responsiveness. A comparative analysis of the immunological phenotype linked to the GRGs associated with the extremes of susceptibility, i.e., *CCL3L1*^{low}*CCR5*^{det} and *CCL3L1*^{high}*CCR5*^{non-det}, could provide key insights into the immune correlates of an effective vaccine. This stems from several vaccine studies in simian models showing that *CCR5* ligand production is a true predictor of protection and animals that produce higher levels of chemokines pre-vaccination exhibit greater protection (20–22).

Finally, and of broader significance, 5% of the human genome contains duplicated sequences enriched for genes involved in immunity (*I*), and some of these genes have dosage effects. Thus, the present findings provide both a precedent and a framework for elucidating their relationship to human diseases.

References and Notes

1. J. A. Bailey *et al.*, *Science* **297**, 1003 (2002).
2. J. R. Townson, L. F. Barcellos, R. J. Nibbs, *Eur J Immunol* **32**, 3016 (2002).
3. P. Menten, A. Wuyts, J. Van Damme, *Cytokine Growth Factor Rev* **13**, 455 (2002).
4. Materials and methods are available as supporting material on *Science Online*.
5. J. W. Mellors *et al.*, *Ann Intern Med* **126**, 946 (1997).
6. L. Wu *et al.*, *J Exp Med* **185**, 1681 (1997).
7. D. Zagury *et al.*, *Proc Natl Acad Sci U S A* **95**, 3857 (1998).
8. A. Garzino-Demo *et al.*, *Proc Natl Acad Sci U S A* **96**, 11986 (1999).
9. J. Reynes, V. Baillat, P. Portales, J. Clot, P. Corbeau, *J Acquir Immune Defic Syndr* **34**, 114 (2003).
10. H. Ullum *et al.*, *J Infect Dis* **177**, 331 (1998).
11. W. A. Paxton *et al.*, *J Infect Dis* **183**, 1678 (2001).
12. J. Tang, R. A. Kaslow, *Aids* **17 Suppl 4**, S51 (2003).
13. M. P. Martin *et al.*, *Science* **282**, 1907 (1998).
14. E. Gonzalez *et al.*, *Proc Natl Acad Sci U S A* **96**, 12004 (1999).
15. A. Mangano *et al.*, *J Infect Dis* **183**, 1574 (2001).
16. S. Mummidi *et al.*, *J Biol Chem* **275**, 18946 (2000).
17. J. R. Salkowitz *et al.*, *Clin Immunol* **108**, 234 (2003).
18. R. V. Samonte, E. E. Eichler, *Nat Rev Genet* **3**, 65 (2002).
19. D. L. Weed, *Hematol Oncol Clin North Am* **14**, 797 (2000).
20. A. L. DeVico, R. C. Gallo, *Nat Rev Microbiol* **2**, 401 (2004).
21. J. L. Heeney *et al.*, *Proc Natl Acad Sci U S A* **95**, 10803 (1998).
22. R. K. Ahmed *et al.*, *Clin Exp Immunol* **129**, 11 (2002).
23. We thank the Board member and reviewers, including the statistical referee, for critically reviewing various aspects

of this work and for very valuable suggestions; G. Crawford, B. Kasinath, G. Nabel, J. Burns, B. Cherniak, members of the Infectious Diseases division for helpful discussions and critical reading of the manuscript; E. Fattig and M. Hildebrand for technical assistance; N. Chopra and J. Sharron for graphic work; and A. S. Ahuja for forbearance. The Henry M. Jackson Foundation and the Military HIV Program, Walter Reed Army Institute of Research contributed support for the WHMC patient cohort as part of the Tri-Service HIV Program. Supported by the Veterans Administration Center on AIDS and HIV-1 infection, and grants from NIH (AI046326, AI043279, MH069270) (S.K.A.). S.K.A. is a recipient of the Elizabeth Glaser Scientist Award and the Burroughs Wellcome Clinical Scientist Award in Translational Research. Because of space constraints we regret our inability to cite additional excellent work.

Supporting Online Materials

www.sciencemag.org/cgi/content/full/1101160DC1

Materials and Methods

SOM Text

Figs S1 to S16

Tables S1 to S7

References and Notes

07 June 2004; accepted 22 December 2004

Published online 06 January 2005; 10.1126/science.1101160

Include this information when citing this paper

Fig. 1. Distribution of *CCL3L1* gene-containing segmental duplications in human populations and *CCL3L* ortholog(s) in chimpanzees (SOM section 1). The cloning and characterization of the chimpanzee orthologs is described in Fig. S3. **(A)** The human populations are labeled below the figure, and their geographic affiliations are shown above it. *CCL3L1* copy number was determined by real-time Taqman PCR assays (SOM sections 2.1 and 4.2). **(B)** Distribution of *CCL3L1* copy numbers in African and non-African human populations and chimpanzee orthologs. The mean, variance, standard deviation (SD), median, and interquartile range (IQR) of the copy numbers are shown here and in Fig. S1B. **(C)** Cumulative frequency curves of the distribution of *CCL3L1* copies in human populations from different geographic regions and chimpanzee. The order of the abbreviations (geographic regions shown in panel A and chimpanzee (CH)) matches the order of the cumulative frequency curves from left to right.

Fig. 2. *CCL3L1* dose and risk of acquiring HIV-1. **(A to D)** Histograms and the cubic-spline smoothed frequency curves (insets) show that the distribution of the *CCL3L1* copy numbers (*x* axis) in HIV⁺ (red bars or red lines in inset)

versus HIV⁻ (open bars or black line in inset) individuals is markedly different (χ^2 and *P* values above insets; *n* = number of individuals in each group). Vertical green arrow (copy number at which the HIV⁺/HIV⁻ ratio switched from >1 to ≤ 1) indicates the switch point (SOM section 5.1). The cohort of Argentinean children is comprised of children exposed perinatally to HIV (4). The HIV⁺ adults from the indicated ethnic/racial groups (noted on the right) are from the Wilford Hall Medical Center (WHMC) cohort (14) and are compared to a control group from the general population that is matched for ethnicity/race (4). (E to H) Risk of acquiring HIV relative to the population-specific median (horizontal arrow; odds ratio (OR) = 1) was determined by multivariate logistic regression analyses. *, Jewell correction (4); #, *CCL3L1* gene copy number; CI, confidence interval; *P*, significance value. (I) Distribution of *CCL3L1* copies in the indicated subject groups (Poisson means and exact 95% CI). #, group number. Arg., Argentinean. SOM section 1.1 provides details of these study groups. Note, in the HIV⁻ WHMC cohort, as HAs were categorized with EAs, they were placed within a single group (WHMC EA + HA) and compared to subjects from the HIV⁺ WHMC cohort that are matched for ethnicity/race. (J) Results of Poisson and logistic regression models in the study groups indicated in panel I (#) for the association between *CCL3L1* copies and risk of acquiring HIV infection were comparable (table S1).

Fig. 3. Disease-influencing and functional phenotypic effects associated with number of *CCL3L1* copies. (A and B) Kaplan–Meier (KM) survival curves of the development of AIDS in (A) AAs and (B) EAs from the adult WHMC HIV⁺ cohort who possess a *CCL3L1* copy number equal to, or lower than the population-specific median (copy numbers noted adjacent to KM curves). Note, as the population-specific median number of *CCL3L1* copies was 3 and 4 in HIV⁺ and HIV⁻ AAs, respectively, these two copy numbers were used as the reference genetic strata in A; the reference group in EAs is 2 copies. *P* and relative hazard (RH) below the KM curves were determined by Cox proportional hazard models. Overall log-rank significance values and 95% CI for the RHs are shown in fig. S5. (C) Relationship between number of *CCL3L1* copies and percentage of CD4⁺/CCR5⁺ cells in unstimulated (open bars) or anti-CD3/CD28-stimulated peripheral blood mononuclear cells (black bars). Numbers inside the bars denote the number of individual blood samples studied with the indicated copy numbers. K-W *P*, overall Kruskal–Wallis test *P* value. Vertically oriented numbers indicate *P* values by the Mann–Whitney test for comparison of possession of 0–2 versus 3–4 or 5–7 *CCL3L1* copies within each experimental condition. (D to F) Second-order polynomial regression curves show (D) CCL3/CCL3L1 concentrations in supernatants of freshly isolated peripheral blood mononuclear cells (for units, see (4); *n* = no. of

individuals), (E) baseline log viral RNA (viral set point), and (F) monthly CD4⁺ T cell loss have a threshold-type association with *CCL3L1* copies (SOM section 4.5–4.7). D and E depict medians (± 1.7 standard deviation of medians), and F depicts 95% CI around the point estimates of the regression coefficients obtained by the General Estimating Equations (GEE) method (4). *P* linear and quadratic (quad) indicate significance values for the linear and quadratic terms in the polynomial regression equation, respectively. (G to L) KM curves of the development of AIDS in HIV⁺ AAs (red) and EAs (green) who possess a similar or dissimilar *CCL3L1* copy number. The disease-influencing effects associated with possession of (G and H) median, (I) half-median, and (J) low/null *CCL3L1* doses were similar in EAs and AAs. However, the disease-influencing effects of possession of (K) two copies in AAs (half-median dose in HIV⁻ AAs) and EAs (median dose) or (L) three copies in AAs (median in HIV⁺ AAs) and one copy in EAs (half-median in EAs) were not equivalent (see also note in A regarding differences in median copy numbers in HIV⁻ and HIV⁺ AAs). Numbers adjacent to the population designators AA and EA indicate the number of copies (e.g., AA4 implies four copies in AAs). *P* values indicate significance value by log-rank test. =, >, or < signifies the direction of the associated effects. (M and N) Direction and magnitude of the rate of change in CD3⁺, CD4⁺ and CD8⁺ T cell counts are similar in HIV⁺ EAs and AAs who possess a *CCL3L1* copy number equal to, or lower than the population-specific median (error bars indicate 95% CI; table S2). (O and P) Results of discrete-time Markov modeling of the evolution of changes in the frequency distribution of *CCL3L1* copy numbers in infinite-sized AA and EA cohorts over 15 years (SOM section 4.8). Numbers adjacent to the curves indicate *CCL3L1* copy numbers. (Q) Schema of phenotypic equivalency of the risk of acquiring HIV and disease-influencing effects of population-specific *CCL3L1* doses in EAs and AAs.

Fig. 4. Risk of acquiring HIV and disease-influencing effects associated with variations in *CCL3L1* and/or *CCR5*. (A) Genetic stratification system (SOM section 3). In each population (pop^l), *CCL3L1* dose and *CCR5* genotypes were dichotomized based on whether they were associated with an accelerated disease course (table S3 to S5). *CCL3L1*^{low} and *CCL3L1*^{high} denote copy numbers < or ≥ population-specific median, respectively (table S3). *CCR5*^{det} and *CCR5*^{non-det} denote population-specific, disease-accelerating, i.e., detrimental (*det*, indicates detrimental), or non-detrimental *CCR5* genotypes, respectively (table S4). Compared with possession of *CCL3L1*^{high} or *CCR5*^{non-det}, *CCL3L1*^{low} or *CCR5*^{det} was associated with an accelerated disease course (fig. S10). These dichotomized compound genotypes were used to stratify the cohort further into four mutually exclusive GRGs, which reflected (i) the independent disease-

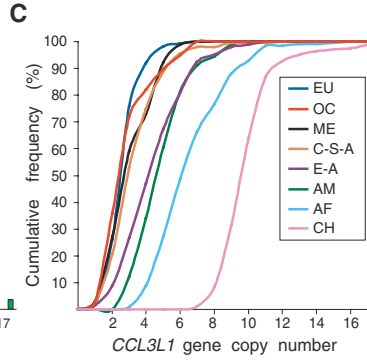
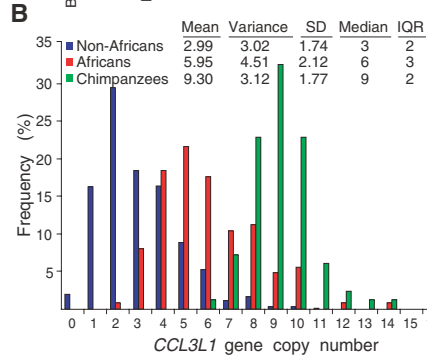
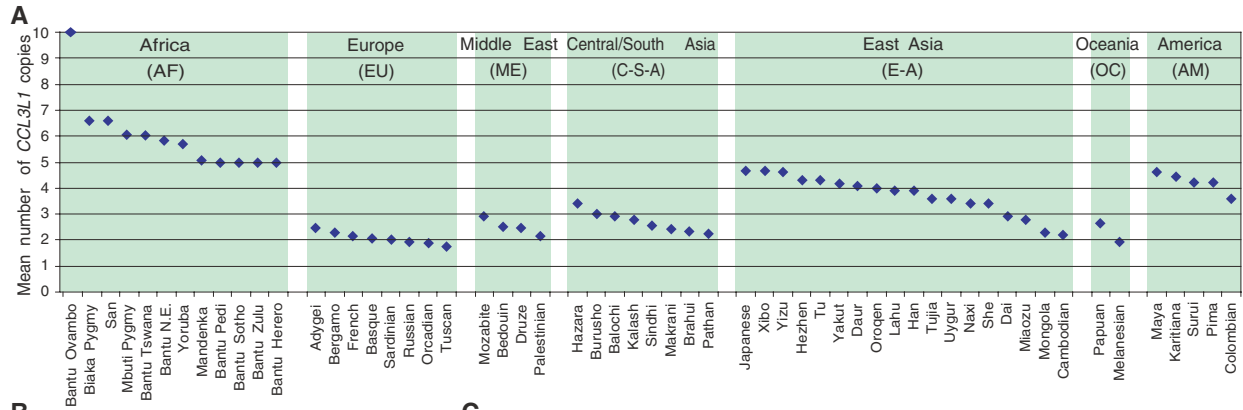
accelerating effects associated with population-specific low *CCL3L1* gene doses (*CCL3L1*^{low}*CCR5*^{non-det}; orange) or detrimental *CCR5* genotypes (*CCL3L1*^{high}*CCR5*^{det}; green); or (ii) their combined effects (*CCL3L1*^{low}*CCR5*^{det}; red), all relative to *CCL3L1*^{high}*CCR5*^{non-det} (blue). This color code is used in the rest of the panels to indicate the four *CCL3L1/CCR5* GRGs. **(B)** CD4⁺ and **(C)** CD8⁺ T cell changes associated with the GRGs are depicted as 95% CI around the point estimates of the regression coefficients obtained by the GEE method (4). **(D)** Baseline log viral RNA (viral set point; median (±1.7 standard deviation of the median)) associated with the GRGs. *P* values reflect significance values for differences between *CCL3L1*^{high}*CCR5*^{non-det} and *CCL3L1*^{low}*CCR5*^{det} by Student's *t*-test in B and C and the Mann–Whitney test in D. **(E and F)** KM curves of the development of AIDS in EAs and AAs from the entire (E) or seroconverting portion (F) of the HIV⁺ adult cohort after stratifying for the GRGs. Inset, pie-charts depicting frequency distribution of the GRGs. **(G)** Proportions of individuals within each GRG that developed AIDS. **(H and I)** Association of indicated GRGs and risk of acquiring HIV infection in (H) adults or (I) children exposed perinatally to HIV. Note, ORs are lowest in GRGs that lack *CCL3L1*^{low} (green). **(J and K)** Changes in the frequency distributions of the GRGs and test of linear trend for individuals with varying follow-up times. **(L)** Differences in the frequency distribution of GRGs between HIV⁺ and HIV⁻ adults. In H and J, to ensure appropriate ethnic/racial matching for the comparisons of the frequency distributions between HIV⁺ and HIV⁻ individuals, these analyses are for the EA, AA and HA portions of the infected adult cohort (tables S3 and S4; tables also show the genotypes used for the pediatric cohort in I) (4).

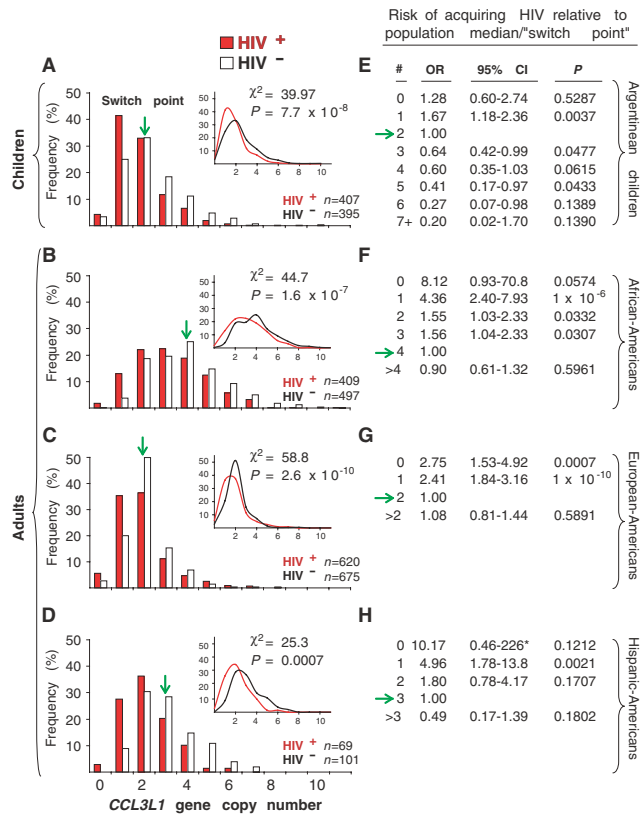
Fig. 5. Attributable fractions of *CCL3L1/CCR5* GRGs for risk of acquiring HIV (vertical, mother-to-child; horizontal, adult-to-adult) and rate of disease progression relative to *CCL3L1*^{high}*CCR5*^{non-det} in the indicated clinical settings. Vertical bars indicate the point estimate, whereas error bars represent the 95% CI around the point estimate of the AF.

Table 1. Risk of AIDS-defining illness with *CCL3L1/CCR5* GRG's. The reference GRG for statistical analysis is *CCL3L1^{high}CCR5^{non-det}* (RH = 1). The AIDS defining illnesses with sufficient events for statistical analyses recorded in the adult HIV⁺ cohort are shown. *CMV, Cytomegalovirus;

HAD, HIV-associated dementia; MAC, *Mycobacterium avium* complex; PCP, *Pneumocystis carinii* pneumonia; PML, Progressive multifocal leukoencephalopathy; N, number of individuals with the indicated AIDS-defining illness; values in red indicate significant association.

AIDS-defining Illness*	N	<i>CCL3L1^{high}CCR5^{det}</i>			<i>CCL3L1^{low}CCR5^{non-det}</i>			<i>CCL3L1^{low}CCR5^{det}</i>		
		RH	95% CI	<i>P</i>	RH	95% CI	<i>P</i>	RH	95% CI	<i>P</i>
CMV infection	100	1.53	0.71-3.30	0.278	1.60	1.00-2.58	0.051	6.21	3.63-10.63	2.7x10 ⁻¹¹
Cryptococcosis	33	3.27	0.98-10.87	0.053	2.46	1.00-6.02	0.048	8.11	2.93-22.46	5.6x10 ⁻⁵
Cryptosporidiosis	24	1.21	0.27-5.47	0.802	1.21	0.49-3.00	0.686	1.63	0.36-7.37	0.526
HAD	54	2.05	0.82-5.13	0.126	1.65	0.87-3.11	0.124	3.18	1.33-7.60	0.009
Herpes simplex	26	1.78	0.50-6.41	0.375	1.22	0.49-3.04	0.668	1.66	0.36-7.53	0.513
Histoplasmosis	20	3.32	0.83-13.30	0.090	2.81	1.02-7.74	0.045	1.56	0.19-13.01	0.682
Kaposi sarcoma	74	1.76	0.76-4.05	0.186	1.66	0.96-2.86	0.069	3.86	1.90-7.83	2.0x10 ⁻⁴
Lymphoma	37	2.87	1.10-7.48	0.031	1.42	0.66-3.08	0.369	3.38	1.21-9.43	0.020
MAC	92	2.22	1.09-4.55	0.029	1.73	1.05-2.87	0.032	5.13	2.79-9.45	1.5x10 ⁻⁷
PCP	196	2.13	1.33-3.42	0.002	1.71	1.22-2.39	0.002	2.95	1.84-4.75	7.8x10 ⁻⁶
PML	18	1.72	0.36-8.10	0.494	1.27	0.44-3.67	0.657	2.41	0.51-11.43	0.268
Toxoplasmosis	27	1.49	0.32-6.91	0.610	1.69	0.67-4.25	0.268	5.34	1.77-16.07	0.003





I

#	HIV ⁺ subjects		HIV ⁻ subjects	
	Subjects	CCL3L1 gene copies	Subjects	CCL3L1 gene copies
		Mean (95% CI)		Mean (95% CI)
1.	Arg. children	1.85 (1.72-1.99)	Arg. children	2.41 (2.26-2.57)
2.	WHMC EA+H A	1.88 (1.74-1.99)	WHMC EA+H A	2.10 (1.99-2.22)
3.	WHMC AA	3.26 (3.08-3.44)	WHMC AA	3.79 (3.60-3.98)
4.	WHMC EA	1.86 (1.75-1.97)	Non-WHMC EA	2.13 (2.02-2.24)
5.	WHMC AA	3.26 (3.08-3.44)	Non-WHMC AA	3.90 (3.73-4.08)
6.	WHMC HA	2.06 (1.73-2.43)	Non-WHMC HA	3.08 (2.75-3.44)

J

#	Poisson regression			Logistic regression		
	RR	95% CI	P	OR	95% CI	P
1.	0.77	0.70-0.85	6.7×10^{-7}	0.71	0.63-0.80	6.7×10^{-9}
2.	0.90	0.83-0.97	0.0059	0.83	0.75-0.92	0.0005
3.	0.86	0.80-0.93	6.6×10^{-5}	0.81	0.74-0.88	3.3×10^{-6}
4.	0.87	0.81-0.95	0.0007	0.80	0.72-0.89	2.2×10^{-5}
5.	0.84	0.78-0.90	4.7×10^{-7}	0.78	0.72-0.85	8.4×10^{-9}
6.	0.67	0.55-0.82	8.5×10^{-5}	0.49	0.36-0.67	9.0×10^{-6}

