

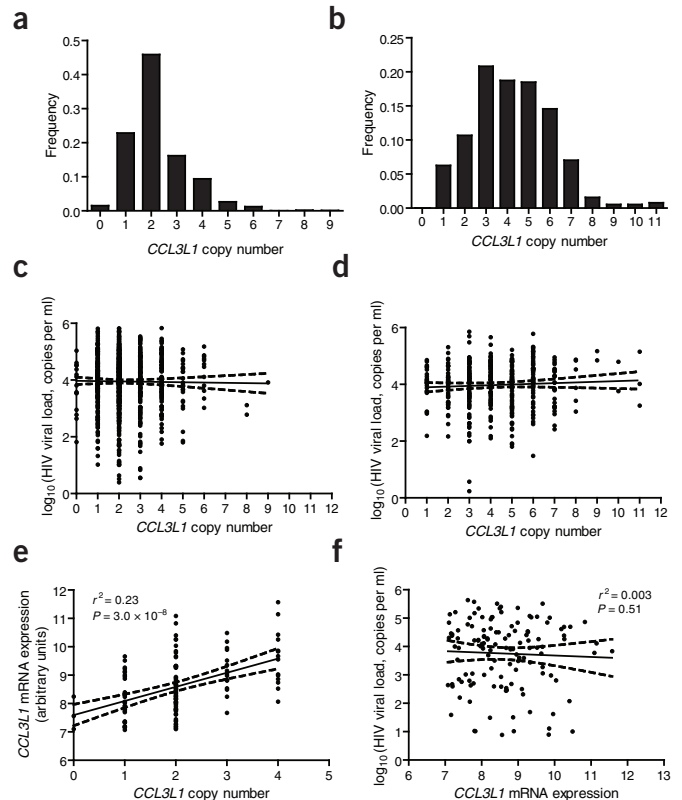
## CCL3L1 and HIV/AIDS susceptibility

### To the Editor:

We read with interest the recent article in *Nature Medicine* describing the influence of variation in *CCL3L1* copy number and *CCR5* genotype on immune recovery during highly active antiretroviral therapy (HAART) in HIV-1-infected individuals<sup>1</sup>. The chemotactic cytokine *CCL3L1* (encoding the macrophage inflammatory protein-1 $\alpha$ P (MIP-1 $\alpha$ P) protein) is a potent ligand for the HIV-1 co-receptor CCR5, which is essential for viral entry into human host cells<sup>2</sup>. The recent study is part of a series that began in 2005 with a paper reporting effects of *CCL3L1* copy number variation on HIV-1 acquisition, viral load and disease progression<sup>3</sup>, followed by several publications investigating clinically correlated phenotypes in a largely overlapping set of HIV-positive individuals<sup>1,4,5</sup>.

Although these studies seem to generate considerable independent support for a role of *CCL3L1* in viral control, many of the traits considered are at least partially correlated, and the studies include largely overlapping samples and presumably *CCL3L1* assay data. For these reasons, we sought to reevaluate a core set of associations related to the effect of *CCL3L1* copy number on viral control in a large group of HIV-infected individuals with known date of seroconversion enrolled in one of the nine cohorts of the Euro-CHAVI Consortium<sup>6</sup> ( $n = 1,042$ ), in an African-American cohort from the TACC ( $n = 277$ ) or in the MACS ( $n = 451$  HIV-positive,  $n = 195$  high-risk seronegative (HRSN)). We assayed for *CCL3L1* copy number using a previously described method<sup>3</sup> (Supplementary Methods). A total of 1,855 subjects were successfully genotyped. Distributions of *CCL3L1* copy numbers in individuals of European or African ancestry were similar to those reported elsewhere, with a median copy number of 2 or 4 in individuals of primarily European (range 0–9) or African (range 1–11) descent, respectively (Fig. 1a,b and Supplementary Figs. 1–4)<sup>1,3,4,7</sup>.

We then tested for association of *CCL3L1* copy number with HIV viral load at set point by linear regression after stratifying according to ethnicity and correcting for known covariates (gender, age at seroconversion and ancestry as determined by a principal components method described previously<sup>8</sup>), and found no evidence of association (European,  $P = 0.14$ ; African,  $P = 0.27$ ) (Fig. 1c,d). Dividing the sample into the previously described “high-risk” (*CCL3L1*<sup>low</sup>) and “low-risk” (*CCL3L1*<sup>high</sup>) genotype groups (where high risk versus low risk is defined as having copy number below versus equal to or above the population median, respectively)<sup>3</sup>, we again found no evidence of association, either within each population (European,  $P = 0.10$ ; African,  $P = 0.41$ ) or in the combined sample ( $P = 0.35$ ) (Table 1). Furthermore, a model including known functional polymorphisms in the *CCR5* receptor (*CCR5* $\Delta$ 32, *CCR5*\*HHE) in a subset of  $n = 820$  individuals of European descent for which *CCR5* effects had been tested previously (J.F., D. Ge, K.V.S., S.C., B. Ledergerber *et al.*, unpublished data) showed that although the *CCR5* polymorphisms were strongly associated with viral load (*CCR5* $\Delta$ 32:  $\beta = -0.29 \pm 0.08$  log RNA copies,  $P = 0.001$ ; *CCR5*\*HHE:  $\beta = 0.14 \pm 0.05$  log RNA copies,  $P = 0.005$ ), there remained no appreciable effect of *CCL3L1* copy number (copy number,  $P = 0.24$ ; genotype risk group,  $P = 0.12$ ).



**Figure 1** *CCL3L1* copy number estimates and their relationship to HIV-1 viral load and *CCL3L1* mRNA expression. (a,b) Distribution of *CCL3L1* copy number in HIV-infected individuals of recent European (a) or African (b) ancestry. (c,d) Relationship between HIV viral load at set point and *CCL3L1* copy number among subjects of recent European (c) or African (d) ancestry. Linear regression of HIV viral load at set point on *CCL3L1* copy number showed no significant effect of *CCL3L1* dose (European:  $r^2 = 0.0006$ ,  $P = 0.14$ ; African:  $r^2 = 0.0022$ ,  $P = 0.27$ ). (e,f) Relationship among *CCL3L1* copy number, *CCL3L1* mRNA expression and viral load at set point.

Next, we tested whether *CCL3L1* variation influences disease progression. We used both a quantitative measure of progression introduced previously<sup>6</sup> (consisting of measured or estimated time to CD4<sup>+</sup> cell count <350 per mm<sup>3</sup> or initiation of antiretroviral therapy; Supplementary Methods) and a simple case-to-control comparison of progressors versus nonprogressors (defined as progression to CD4<sup>+</sup> cell count <350 per mm<sup>3</sup> or antiretroviral therapy within 10 years since seroconversion versus no progression within 10 years). Finally, we tested for an effect of *CCL3L1*<sup>low</sup> versus *CCL3L1*<sup>high</sup> group assignment on these measures as well as on progression to AIDS 1987, AIDS 1993 or AIDS-related death using a Cox proportional hazards model. Neither *CCL3L1* copy number nor *CCL3L1*<sup>low</sup> versus *CCL3L1*<sup>high</sup> genotype group assignment was associated with disease progression under any of these models

**Table 1** Results of statistical tests for association of *CCL3L1* copy number or genotype risk group (GRG) status with HIV infection risk or HIV-related outcomes.

Continuous traits	<i>n</i>	$\beta$	$r^2$	<i>P</i> value
Set point versus copy number (EUR)	1138	-0.0342	0.0006	0.14
Set point versus copy number (AFR)	366	0.0244	0.0022	0.27
Set point versus <i>CCL3L1</i> <sup>high</sup> or <i>CCL3L1</i> <sup>low</sup> (EUR)	1138	-0.1022	0.0019	0.10
Set point versus <i>CCL3L1</i> <sup>high</sup> or <i>CCL3L1</i> <sup>low</sup> (AFR)	366	0.0741	0.0021	0.41
Set point versus <i>CCL3L1</i> <sup>high</sup> or <i>CCL3L1</i> <sup>low</sup> (combined)	1504	-0.0395	0.0006	0.42
Time to progression versus copy number (EUR)	682	6.64	0.0002	0.90
Time to progression versus <i>CCL3L1</i> <sup>high</sup> or <i>CCL3L1</i> <sup>low</sup> (EUR)	682	62.5	0.0001	0.65
Binary traits	<i>n</i> (progressor)	<i>n</i> (nonprogressor)	Odds ratio	<i>P</i> value
Progressor or nonprogressor versus copy number (EUR)	611	71	1.14	0.33
Progressor/nonprogressor versus <i>CCL3L1</i> <sup>high</sup> or <i>CCL3L1</i> <sup>low</sup> (EUR)	611	71	1.25	0.46
Survival analysis	<i>n</i>	Hazard ratio	95% CI	<i>P</i> value
Time to progression versus <i>CCL3L1</i> <sup>high</sup> or <i>CCL3L1</i> <sup>low</sup> (EUR)	744	0.95 +/- 0.09	0.78–1.15	0.59
Tests for association with infection status	<i>n</i> (HIV <sup>+</sup> )	<i>n</i> (HRSN)	Odds ratio (95% CI)	<i>P</i> value
HIV infection status versus copy number	451	195	0.86–1.08	0.53
HIV infection status versus <i>CCL3L1</i> <sup>high</sup> or <i>CCL3L1</i> <sup>low</sup>	451	195	0.52–1.13	0.18

EUR, individuals of recent European ancestry; AFR, individuals of recent African ancestry (African Europeans and African Americans); 95% CI, 95% confidence interval.  $\beta$ , odds ratios, hazard ratios and *P* values are reported for the *CCL3L1* copy number or GRG term after adjusting for gender, age at seroconversion and population structure.  $r^2$  values represent the fraction of variation explained by the *CCL3L1* copy number or GRG term before correction for other covariates.

( $P > 0.1$  for all tests) (Table 1).

We then tested whether *CCL3L1* copy number is associated with risk of HIV infection by comparing the copy number distributions in HIV-infected individuals compared with individuals who were judged to be unusually exposed to HIV but remain uninfected (called HSRN). Using samples from the MACS cohort, we compared 451 HIV-positive to 195 HRSN individuals. This comparison was well powered to detect effects of *CCL3L1* copy number on risk of infection through mucosal exposure (the principal model of transmission in this cohort; Supplementary Table 1). We found no association between infection status and either copy number ( $P = 0.53$ ) or genotype risk group ( $P = 0.18$ ) (Table 1). In the same sample, *CCR5Δ32* homozygosity was strongly associated with reduced risk of infection (the *CCR5Δ32*-homozygous genotype frequency was 4.9% in exposed uninfected versus 0% in infected individuals,  $P = 3.5 \times 10^{-6}$ ). Of note is the enrichment of *CCR5Δ32* homozygotes in the HRSN sample (4.9% versus an estimated 1% in unselected individuals of European descent)<sup>9</sup>, indicating that the effective exposure in the HRSN cohort was very high and therefore that this cohort should provide sufficient power to detect additional genetic risk factors of reasonable effect size. Notably, we also found no effect of *CCL3L1* copy number on infection risk after stratifying according to *CCR5Δ32* genotype.

We investigated whether *CCL3L1* copy number influences *CCL3L1* messenger RNA expression in CD4<sup>+</sup> T lymphocytes from 122 HIV-positive individuals who had not yet initiated antiretroviral therapy, using the Illumina WG-6 v3 expression array (Supplementary Methods), and found a strong and linear increase of *CCL3L1* mRNA levels with copy number ( $r^2 = 0.23$ ,  $P = 3.0 \times 10^{-8}$ , Fig. 1e). In the same samples, however, *CCL3L1* mRNA expression itself showed no correlation with HIV set point ( $r^2 = 0.003$ ,  $P = 0.51$ , Fig. 1f).

These observations raise the question of why earlier studies reported positive associations that we could not replicate here. As a possible explanation, we note that measurement of *CCL3L1* copy number variation seems to be highly susceptible to systematic biases related to the preparation and quality of DNA samples. We observed that batch

differences in input DNA amounts between cases and controls can lead to biased copy number estimates by the real-time PCR method used here, and, in fact, we found an apparently considerable association in the direction opposite to that previously reported<sup>3</sup> (with higher copy number among HIV<sup>+</sup> cases compared with controls) before diluting DNA samples into an appropriate range (Supplementary Table 2). Additionally, we compared the results of various assays (the real-time PCR-based assay used here and in the previous reports<sup>1,3,4</sup> and a recently published method based on the paralogue ratio test (PRT)<sup>10,11</sup>) and found that although the results were generally very highly correlated, for one comparison the association statistics from the two assays diverged markedly. Specifically, in a comparison of a small number of HIV-positive and HIV-negative samples from Malawi, copy number estimated by the PRT method showed a strong association with infection status, whereas the real-time PCR-based estimates showed no association; this discrepancy seems to be explained by systematic differences in DNA degradation between case and control samples, in which degradation or shearing of DNA leads to systematic overestimation of copy number by the PRT method specifically (Supplementary Table 3). Among HIV-positive individuals, we did not observe either assay method recording a signal of association for any HIV-related quantitative trait; these tests are both more statistically powerful than the case-to-control comparisons and far less sensitive to any 'batch effects' on the copy number estimation. Although both of the assays described here are liable to various types of systematic biases, we emphasize that differences between cohorts in the distribution of DNA concentrations are presumed to be far more likely, and perhaps expected, compared with differences in DNA storage or degradation, and thus the real-time PCR method will often be expected to produce a false positive association unless input DNA amounts are carefully considered. We therefore suggest that some of the previously reported associations may reflect differences in DNA quality or concentration that systematically increase or decrease the inferred number of copies of *CCL3L1* in cases versus control samples.

In summary, we find the absence of any substantial effect of *CCL3L1* copy number variation on HIV-1 infection, viral load or disease progression. We do, however, show a highly significant ( $P = 3.0 \times 10^{-8}$ ) association between copy number variation and *CCL3L1* mRNA levels, demonstrating that the assays are sufficiently accurate to detect the intermediate biological effects of copy number variation. Although there is some evidence that reduced expression of the CCR5 receptor may aid in viral control and delay progression to AIDS, there is less reason to believe that CCR5 inhibition is protective from infection without complete CCR5 blockade. Other groups have shown that *CCL3L1* mRNA and MIP-1 $\alpha$ P protein are expressed at relatively low levels compared with other CCR5 ligands, with measured serum concentrations of MIP-1 $\alpha$ P protein well below its estimated half-maximal effective concentration, determined on the basis of *ex vivo* assays<sup>7,12</sup>. Indeed, concentration increases to orders of magnitude higher than those reported in both healthy and HIV-infected individuals seem to be required to approach half-maximal occupancy by *CCL3L1*, whereas a reduction in receptor concentration would effectively multiply the affinity of all CCR5 ligands, including the much more abundant CCL5 (RANTES) and others in addition to MIP-1 $\alpha$ P<sup>7</sup>. We should note that these arguments do not apply to the postulated effects of *CCL3L1* that operate independently of direct CCR5 blockade, for example through effects on the expression of innate defense pathways. Such an explanation, however, might also be expected to drive correlations between *CCL3L1* expression and viral control in infected individuals, which we did not observe (Fig. 1d). We point out that a gold standard for copy number determination in this region is yet lacking and that the current techniques are likely to be influenced by other sources of error beyond the systematic ones described here. Despite progress in cataloging sequence and structural variation in the *CCL3L1* region<sup>13</sup>, accurate assessment of the contribution of genetic variation in such a complex region will require the development of more accurate assay methods that provide information about not only gene copy number but also gene content. Finally, we emphasize that these results do not cast any doubt on efforts to develop CCR5 antagonists (that is, MIP-1 $\alpha$ P analogs) as therapeutics for HIV prevention and treatment but merely argue that natural variation in *CCL3L1* gene dose does not seem to have any major effects on the control of HIV-1.

All experiments were approved by the Duke University Institutional Review Board. Informed consent was obtained from all participating subjects.

Note: Supplementary information is available on the Nature Medicine website.

Thomas J Urban<sup>1</sup>, Amy C Weintrob<sup>2,15</sup>, Jacques Fellay<sup>1</sup>, Sara Colombo<sup>3</sup>, Kevin V Shianna<sup>4</sup>, Curtis Gumbs<sup>1</sup>, Margalida Rotger<sup>3</sup>, Kimberly Pelak<sup>1</sup>, Kristen K Dang<sup>1</sup>, Roger Detels<sup>5,14</sup>, Jeremy J Martinson<sup>6,14</sup>,

Stephen J O'Brien<sup>7</sup>, Norman L Letvin<sup>8</sup>, Andrew J McMichael<sup>9</sup>, Barton F Haynes<sup>10</sup>, Mary Carrington<sup>11</sup>, Amalio Telenti<sup>3,13</sup>, Nelson L Michael<sup>12,15</sup> & David B Goldstein<sup>1,13</sup>

<sup>1</sup>Center for Human Genome Variation, Duke Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina, USA. <sup>2</sup>Infectious Disease Clinical Research Program, Walter Reed Army Medical Center, Washington, DC, USA. <sup>3</sup>Institute of Microbiology, University Hospital Center, and University of Lausanne, Lausanne, Switzerland. <sup>4</sup>Duke Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina, USA. <sup>5</sup>Department of Medicine, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California, USA. <sup>6</sup>Department of Infectious Diseases and Microbiology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. <sup>7</sup>Laboratory of Genomic Diversity, National Cancer Institute at Frederick, Frederick, Maryland, USA. <sup>8</sup>Division of Viral Pathogenesis, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA. <sup>9</sup>Medical Research Council Human Immunology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, UK. <sup>10</sup>Duke Human Vaccine Institute, Duke University, Durham, North Carolina, USA. <sup>11</sup>Cancer and Inflammation Program, Laboratory of Experimental Immunology, SAIC-Frederick, Inc., National Cancer Institute at Frederick, Frederick, Maryland, USA. <sup>12</sup>Division of Retrovirology, Walter Reed Army Institute of Research, US Military HIV Research Program, Rockville, Maryland, USA. <sup>13</sup>Center for HIV/AIDS Vaccine Immunology (CHAVI) and the Euro-CHAVI consortia. <sup>14</sup>Multicenter AIDS Cohort Study (MACS). <sup>15</sup>Tri-Service AIDS Clinical Consortium (TACC) of the Infectious Disease Clinical Research Program. e-mail: d.goldstein@duke.edu

#### ACKNOWLEDGEMENTS

Funding was provided by the US National Institutes of Health—funded CHAVI. J.F. is supported by the Swiss Foundation for Grants in Biology and Medicine, and A.T. is supported by Infectigen and the Swiss National Science Foundation. This project has been funded in whole or in part with federal funds from the US National Cancer Institute, National Institutes of Health, under contract N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US government. This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research.

- Ahuja, S.K. *et al. Nat. Med.* **14**, 413–420 (2008).
- Townson, J.R., Barcellos, L.F. & Nibbs, R.J. *Eur. J. Immunol.* **32**, 3016–3026 (2002).
- Gonzalez, E. *et al. Science* **307**, 1434–1440 (2005).
- Dolan, M.J. *et al. Nat. Immunol.* **8**, 1324–1336 (2007).
- Kulkarni, H. *et al. PLoS One* **3**, e3165 (2008).
- Fellay, J. *et al. Science* **317**, 944–947 (2007).
- Shao, W. *et al. Genes Immun.* **8**, 224–231 (2007).
- Price, A.L. *et al. Nat. Genet.* **38**, 904–909 (2006).
- Stephens, J.C. *et al. Am. J. Hum. Genet.* **62**, 1507–1515 (1998).
- Armour, J.A. *et al. Nucleic Acids Res.* **35**, e19 (2007).
- Walker, S., Janyakhantikul, S. & Armour, J.A. *Genomics* **93**, 98–103 (2009).
- Menten, P. *et al. J. Clin. Invest.* **104**, R1–R5 (1999).
- Paximadis, M., Mohanlal, N., Gray, G.E., Kuhn, L. & Tiemessen, C.T. *Int. J. Immunogenet.* **36**, 21–32 (2009).

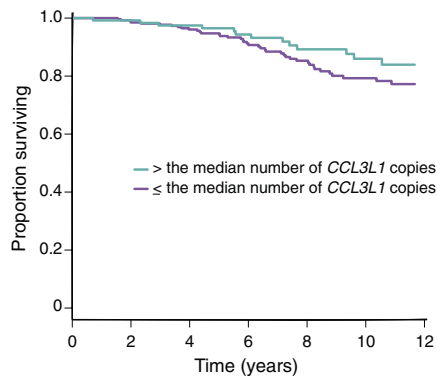
#### To the Editor:

A selective advantage against infectious diseases such as HIV/AIDS is associated with differences in the genes relevant to immunity and virus replication. The CC chemokine receptor-5 (CCR5), the principal co-receptor for HIV, and its chemokine ligands, including CC chemokine ligand-3-like-1 (*CCL3L1*), influence the susceptibility of the CD4<sup>+</sup> target cell to infection<sup>1</sup>. The *CCL3L1* gene is one of several cytokine genes clustered in a region of segmental duplication on chromosome 17q12 (ref. 2). The *CCL3L1* gene copy number per diploid genome varies between people and among persons of various geographical ancestries<sup>2–5</sup>; some people have a partial *CCL3L1* pseudogene as well<sup>2</sup>. Evidence has been presented that increased *CCL3L1* gene copy numbers

can have marked phenotypic consequences for HIV/AIDS pathogenesis as a result of altering gene dosage<sup>3–5</sup>. We have been unable to reproduce these findings using the same methods in a different cohort.

We conducted a population-based study of men enrolled in the Chicago component of the Multicenter AIDS Cohort Study (MACS, **Supplementary Methods**) to estimate the influence of the population-specific *CCL3L1* gene copy number on the risk of HIV infection and/or disease progression and the robustness of immune recovery after the initiation of highly active antiretroviral therapy (HAART). We measured *CCL3L1* gene copy number by real-time PCR assays<sup>3</sup> with modification. For each genomic DNA sample, we determined the *CCL3L1* and *CCR5* gene copy numbers in triplicate (**Supplementary Methods**). We

**Figure 1** HIV disease-influencing effect associated with number of *CCL3L1* gene copies during the pre-HAART era. Kaplan-Meier survival curves are shown for European-American men who possess more than or less than or equal to the population-specific median number of *CCL3L1* copies (two copies). Compared with possession of more than two *CCL3L1* gene copies, men possessing less than or equal to two *CCL3L1* gene copies had a higher risk of AIDS (13 of 124 European-American men compared with 40 of 272 European-American men over the study period, respectively), but the difference is not significant. Men who became infected with HIV during follow-up were at risk for event from the time of their estimated infection, defined as the midpoint between their last HIV antibody-negative and first HIV antibody-positive visits. Men who were infected with HIV at the time they entered the study were treated as censored observations; that is, the time of infection was unknown.



validated standards constructed for absolute quantification of *CCL3L1* gene copy number with sequential dilutions of A431 cell lysate that contains two copies of the *CCR5* gene and two copies of the *CCL3L1* gene per diploid genome<sup>6</sup>. The measured distribution of *CCL3L1* gene copy numbers peaked near biologically interpretable integer values and was rounded to the nearest integer value in all calculations (**Supplementary Fig. 1a,b**). The frequency distribution of *CCL3L1* gene copy numbers varied across populations with median copy numbers of 2 (interquartile range (IQR) defined as the twenty-fifth to seventy-fifth percentiles, 2 to 3) in the sample of 740 European-American men, 4 (IQR 2 to 4) in the 53 white or black Hispanic-American men and 4 (IQR 3 to 6) in the 224 African-American men, similar to those reported by others<sup>3–5</sup>. The difference between the European-American and African-American men is highly significant (Mann-Whitney *U* test,  $P < 2.2 \times 10^{-16}$ ). The Hispanic-American men differed from both the European-American and the African-American men (Mann-Whitney *U* test,  $P = 1.49 \times 10^{-7}$  and  $P = 6.0 \times 10^{-4}$ , respectively), but there was little power to distinguish between the white and black Hispanic-American men.

To determine the influence of *CCL3L1* gene copy number on the risk of acquiring HIV, we assessed the difference in the frequency distribution between HIV-infected ( $n = 580$ ) and uninfected ( $n = 437$ ) men. We randomly distributed the genomic DNA for these groups of men across and among the microtiter plates, and the sample values were reproducible between measurements. We found no evidence that the number of *CCL3L1* gene copies is conditional on the HIV infection status within each geographic ancestral population (median 2 (IQR 2 to 3) for both groups in European-Americans, Mann-Whitney *U* test,  $P = 0.95$ ; median 4.5 (IQR 3 to 6) in HIV-infected European American men compared to median 4 (IQR 3 to 5) in uninfected African American men,  $P = 0.11$ ) (**Supplementary Fig. 1**). The HIV infection status did not differ significantly in men carrying *CCL3L1* gene copy numbers lower than the population-specific median (95% confidence interval (95% CI) on odds ratio = 0.89 [0.66, 1.19] Fisher's exact  $P = 0.43$ )<sup>7–9</sup>. Fitting a generalized linear model, including a score indicating risk of infection based on behavior<sup>10</sup> and *CCL3L1* gene copy number relative to each population-specific median (designated *CCL3L1*<sup>high</sup> and *CCL3L1*<sup>low</sup>), only the risk score is highly significant ( $P = 2 \times 10^{-8}$ ); *CCL3L1* gene copy number is not ( $P = 0.87$ ).

To test for association between *CCL3L1* gene copy number and the risk of progression to AIDS, we graphically assessed the development of an AIDS-defining illness before 1996 (the pre-HAART era) by Kaplan-Meier survival curves (**Fig. 1**). Although the lines overlapped, we observed a separation of the curves later in time (**Fig. 1**). Consequently, we examined the association between *CCL3L1* gene copy number and very slow progression to AIDS. We classified men according to whether AIDS occurred within 9 years of infection (that is, the expected median progression time observed before 1996) or whether they remained free of AIDS for at least 12 years in the absence of HAART. The 46 European-American men with incident infection who developed AIDS within 9 years (typical progressors) had a lower *CCL3L1* gene copy number than the 88 European-American men who remained free of AIDS for at least 12 years (slow progressors; **Supplementary Fig. 2**). Whereas both groups of men had a median *CCL3L1* gene copy number of 2, 75% of the typical progressors had 2 *CCL3L1* gene copies or lower, and 75% of the slow progressors had two *CCL3L1* gene copies or higher (Mann-Whitney *U* test,  $P = 0.023$ ; **Supplementary Fig. 2**). This finding is consistent with a previous report<sup>3</sup> and may indicate a real effect of *CCL3L1* gene copy number variation on HIV disease progression or a false-positive association signal that arises from the aspects of analysis common to both studies, such as rounding the estimates derived from real-time PCR assays to determine *CCL3L1* gene copy numbers to the closest integer.

In HIV-infected men, we found no relationship between the *CCL3L1* gene copy number and the level of HIV RNA in plasma after resolution of acute infection (Kendall's rank correlation  $P = 0.5$  and  $P = 0.6$  in European-American men and African-American men, respectively) or the CD4<sup>+</sup> T cell slope (Kendall's rank correlation  $P = 0.08$  and  $P = 0.35$  in European-American men and African-American men, respectively). When we dichotomized the *CCL3L1* gene copy number at the median, neither the CD4<sup>+</sup> T cell slope (median  $-52.2$  cells per  $\mu\text{l}$  per month (IQR  $-106$  to  $-21$ ) against median  $-51.5$  cells per  $\mu\text{l}$  per month (IQR  $-96$  to  $-23$ ); Mann-Whitney *U* test,  $P = 0.88$ ) nor the level of HIV RNA in plasma at set point (median 16,310 copies per ml (IQR 3,446 to 38,490) against median 16,810 copies per ml (IQR 4,364 to 41,330); Mann-Whitney *U* test,  $P = 0.33$ ) differed between the men possessing *CCL3L1* gene copy numbers lower than the median compared to the rest of the population.

We next measured the size and the significance of the interaction between the *CCL3L1* gene copy number and the *CCR5*  $\Delta 32$  protein-inactivating deletion allele (*CCR5* $\Delta 32$ ). Notably, *CCR5* $\Delta 32$  and the seven single nucleotide polymorphisms within the *CCR5* human haplogroup E (HHE) that influences HIV/AIDS susceptibility (that is, the principal detrimental *CCR5* genotype previously reported<sup>3</sup>) are in strong linkage disequilibrium with each other (reported previously for men from the MACS<sup>11</sup>). The men were stratified into groups on the basis of whether they possessed a *CCL3L1* gene copy number higher than or equal to or lower than the population-specific median (*CCL3L1*<sup>high</sup> or *CCL3L1*<sup>low</sup>, respectively) and the presence or absence of *CCR5* $\Delta 32$  (*CCR5*<sup>wild-type</sup> or *CCR5* $\Delta 32$ , not distinguishing between the *CCR5* $\Delta 32/\Delta 32$  and *CCR5* $\Delta 32/\text{wild-type}$  genotypes) (**Table 1**). We analyzed the influence of the *CCL3L1*-*CCR5* genotype on HIV infection by means of Cox proportional hazards models after adjusting for the behavioral risk score<sup>10</sup>. In this analysis, we did not find an HIV infection-influencing effect of *CCR5* $\Delta 32$  (likelihood ratio = 0.18; hazard ratio = 0.75) or *CCL3L1*<sup>low</sup> (likelihood ratio = 0.42; hazard ratio = 0.83). This conclusion remained unchanged when we performed the Cox proportional hazards model analysis with an interaction variable.

Although the *CCR5* $\Delta 32/\Delta 32$  genotype was highly enriched among the uninfected men (95% CI on odds ratio [2.25, 672.19]; Fisher's exact,  $P = 0.0005$  for all men and  $P = 0.007$  for European-American men),

**Table 1** The frequency of *CCL3L1* gene copy numbers relative to the population-specific median and *CCR5* genotypes in European-American, African-American and Hispanic-American men.

<i>CCL3L1</i> gene copies		European-American men		African-American men		Hispanic-American men	
		< median	≥ median	< median	≥ median	< median	≥ median
Uninfected	Total	66	272	33	49	8	9
	<i>CCR5</i> <sup>Δ32/Δ32</sup>	1	9	0	1	0	0
	<i>CCR5</i> <sup>Δ32/wild-type</sup>	19	41	0	1	0	2
	<i>CCR5</i> <sup>wild-type</sup>	42	199	31	45	8	7
HIV-infected	Total	91	311	46	96	18	18
	<i>CCR5</i> <sup>Δ32/Δ32</sup>	0	1	0	0	0	0
	<i>CCR5</i> <sup>Δ32/wild-type</sup>	17	64	4	2	3	0
	<i>CCR5</i> <sup>wild-type</sup>	69	234	40	94	13	18

The totals also include 52 men with undetermined *CCR5* genotype. The number of *CCL3L1* gene copies is higher than or equal to or lower than the population-specific median (that is, *CCL3L1*<sup>high</sup> and *CCL3L1*<sup>low</sup>, respectively). Only European-American men possessing *CCR5*<sup>Δ32/Δ32</sup> or *CCR5*<sup>wild-type</sup> alleles had a significantly different risk for acquiring HIV (Fisher's exact test,  $P = 0.007$  for association between *CCR5* genotype and HIV infection status).

there was no difference between the infection status of men possessing the *CCR5*<sup>Δ32/wild-type</sup> and the *CCR5*<sup>wild-type</sup> genotypes (95% CI on odds ratio [0.7, 1.4]; Fisher's exact,  $P = 0.93$  for all men) (Table 1). The median or mean level of viral RNA in plasma at set point is about half in the men possessing the *CCR5*<sup>Δ32/wild-type</sup> genotype (median 9,520 copies per ml (IQR 1,820 to 28,520)) compared to those men possessing the *CCR5*<sup>wild-type</sup> genotype (median 17,310 copies per ml (IQR 4,647 to 45,740)); Mann-Whitney  $U$  test,  $P = 0.02$ ). When compared to the median, the *CCL3L1* gene copy number had no effect within either of these groups. In particular, among the *CCR5*<sup>Δ32/wild-type</sup>-*CCL3L1*<sup>low</sup> and *CCR5*<sup>Δ32/wild-type</sup>-*CCL3L1*<sup>high</sup> groups, the median level of HIV RNA in plasma at set point was 8,109 copies per ml (IQR 1,208 to 24,020) and 10,180 copies per ml (IQR 1,852–30,520), respectively (Mann-Whitney  $U$  test,  $P = 0.38$ ). Similarly, among the *CCR5*<sup>wild-type</sup>-*CCL3L1*<sup>low</sup> and *CCR5*<sup>wild-type</sup>-*CCL3L1*<sup>high</sup> groups, the median level of HIV RNA in plasma at set point was 15,700 copies per ml (IQR 3,495 to 39,760) and 17,660 copies per ml (IQR 4,964 to 480,400), respectively (Mann-Whitney  $U$  test,  $P = 0.25$ ). Although *CCR5*<sup>Δ32</sup> has an effect on the level of HIV RNA at set point, the effect is independent of the *CCL3L1* gene copy number.

The CD4<sup>+</sup> T cell slope did not differ between the *CCR5*<sup>Δ32/wild-type</sup> (median -56 cells per μl per month (IQR -101 to -24)) and the *CCR5*<sup>wild-type</sup> (median -52 cells per μl per month (IQR -97 to -23)) groups (Mann-Whitney  $U$  test,  $P = 0.62$ ), the *CCR5*<sup>Δ32/wild-type</sup>-*CCL3L1*<sup>low</sup> (median -43 cells per μl per month (IQR -81 to -5)) and *CCR5*<sup>Δ32/wild-type</sup>-*CCL3L1*<sup>high</sup> (median -58 cells per μl per month (IQR -100 to -32)) groups (Mann-Whitney  $U$  test,  $P = 0.27$ ) nor between *CCR5*<sup>wild-type</sup>-*CCL3L1*<sup>low</sup> (median -55 cells per μl per month (IQR -108 to -24)) and *CCR5*<sup>wild-type</sup>-*CCL3L1*<sup>high</sup> (median -51 cells per μl per month (IQR -97 to -23)) groups (Mann-Whitney  $U$  test,  $P = 0.49$ ). We obtained similar results when restricting the analysis to the European-American subpopulation; specifically, we found a lower level of HIV RNA at set point in the *CCR5*<sup>Δ32/wild-type</sup> (median 8,994 copies per ml (1,736 to 28,190)) than the *CCR5*<sup>wild-type</sup> (median 18,380 copies per ml (5,350 to 48,280)) groups (Mann-Whitney  $U$  test,  $P = 0.01$ ) and no difference in the CD4<sup>+</sup> T cell slope between the *CCR5*<sup>Δ32/wild-type</sup> (median -57 cells per μl per month (IQR, -96 to -30)) and *CCR5*<sup>wild-type</sup> (median -57 cells per μl per month (IQR, -104 to -29)) groups (Mann-Whitney  $U$  test,  $P = 0.78$ ). In addition, there was no effect of *CCL3L1* gene copy number on either of these outcomes when restricted to the *CCR5*<sup>wild-type</sup> (Kendall rank correlation,  $P = 0.38$  for the set point and  $P = 0.04$  for the slope) and *CCR5*<sup>Δ32/wild-type</sup> (Kendall rank correlation,  $P = 0.43$  for the set point and  $P = 0.99$  for the slope) grouping.

We next tested whether the *CCL3L1*-*CCR5* genotype influenced the

extent of reconstitution of the CD4<sup>+</sup> T cell count during HAART<sup>4</sup>. We used the square root of the CD4<sup>+</sup> T cell count as the outcome in a repeated-measures regression model, accounting for the correlated measurements contributed by the same person over time. The model included the CD4<sup>+</sup> T cell count and the level of HIV RNA in plasma in the year before the start of HAART with an indicator for whether clinical AIDS had been diagnosed before its initiation. We stratified the model by the CD4<sup>+</sup> T cell count (that is, <350 cells per μl and >350 cells per μl) to determine whether the stage of HIV disease at the time of initiation of HAART modified the effect of the *CCL3L1*-*CCR5* genotype on immune reconstitution<sup>4</sup>.

Using general estimating equations (Supplementary Methods), the mean rate of change in the CD4<sup>+</sup> T cell count for European American men after initiation of HAART showed no distinction between the *CCL3L1*<sup>high</sup> or *CCL3L1*<sup>low</sup> group either within the first 2 years after the start of HAART (coefficient 0.04 per year per *CCL3L1* gene copy number with s.e.m. 0.41), or at any time thereafter (coefficient -0.02 per year per *CCL3L1* gene copy number with s.e.m. 0.07), whether or not the CD4<sup>+</sup> T cell count was stratified at its initiation. There was no difference in the recovery rate for the CD4<sup>+</sup> T cell count indexed from the time of initiation of HAART between the low-risk (*CCL3L1*<sup>high</sup>-*CCR5*<sup>Δ32/wild-type</sup>; mean -0.2 cells per μl per month, s.e.m. 0.6), medium risk (*CCL3L1*<sup>low</sup>-*CCR5*<sup>Δ32/wild-type</sup> or *CCL3L1*<sup>high</sup>-*CCR5*<sup>wild-type</sup>; mean 0.1 cells per μl per month, s.e.m. 0.4), and high-risk (*CCL3L1*<sup>low</sup>-*CCR5*<sup>wild-type</sup>; mean -0.1 cells per μl per month, s.e.m. 1.0) groups. These data confirm the previously reported lack of association between *CCR5*<sup>Δ32/wild-type</sup> and the CD4<sup>+</sup> T cell count after initiation of HAART<sup>12</sup>.

Our findings do not support a relationship between the population-specific *CCL3L1* gene copy number and HIV/AIDS susceptibility or the response to HAART as previously reported<sup>3,4</sup>, even though we had a similar population, used the same method for determining the *CCL3L1* gene copy number, used *CCR5*<sup>Δ32</sup> as a proxy for the *CCR5* promoter polymorphisms<sup>11</sup> and used a study powered to detect the postulated effect (for example, a 62% power to observe the difference of 1 cell per μl per month between the high- and medium-risk groups at  $P = 0.1$ ). Of note, three previous studies<sup>3–5</sup> determined the *CCL3L1* gene copy number in the same HIV-infected ( $n = 1,127$ ) and uninfected ( $n = 1,105$ ) people (that is, individuals enrolled in the Wilford Hall Medical Center Cohort, a component of the United States Military Tri-Service AIDS Clinical Consortium Natural History Cohort Study, with the addition of an uninfected comparison group ( $n = 1,274$ ) and HIV-infected and uninfected Argentinean children ( $n = 802$ )<sup>3</sup> and HIV-infected people enrolled in the Acute Infection and Early Disease Program<sup>4,5</sup> ( $n = 315$  and  $n = 55$ , respectively)). Although we confirm

a nonrandom distribution of *CCL3L1* gene copy numbers in men of different geographical ancestry, possession of a *CCL3L1* gene copy number above or below the population-specific median did not influence the risk of HIV infection or disease progression in the population examined. Accordingly, the failure to replicate does not necessarily indicate a population-specific difference. These results challenge the idea that the *CCL3L1* gene expression phenotype contributes to HIV/AIDS susceptibility.

The Institutional Review Boards of Northwestern University, Howard Brown Health Center, and the Cook County Core Center approved of this study. All participants provided written informed consent.

Note: Supplementary information is available on the Nature Medicine website.

Tanmoy Bhattacharya<sup>1,2</sup>, Jennifer Stanton<sup>3</sup>, Eun-Young Kim<sup>3</sup>, Kevin J Kunstman<sup>3</sup>, John P Phair<sup>3</sup>, Lisa P Jacobson<sup>4</sup> & Steven M Wolinsky<sup>3</sup>

<sup>1</sup>Santa Fe Institute, Santa Fe, New Mexico, USA. <sup>2</sup>Los Alamos National Laboratory, Los Alamos, New Mexico, USA. <sup>3</sup>Division of Infectious Diseases,

The Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA. <sup>4</sup>Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA. e-mail: s-wolinsky@northwestern.edu

#### ACKNOWLEDGMENTS

We thank B. Korber for assistance with the figures. This work is funded by the US National Institutes of Health (AI-35039 to S.M.W. and J.P.P.).

1. Moore, J.P. & Klasse, P.J. *Cell Host Microbe* **2**, 281–283 (2007).
2. Modi, W.S. *Genomics* **83**, 735–738 (2004).
3. Gonzalez, E. *et al. Science* **307**, 1434–1440 (2005).
4. Ahuja, S.K. *et al. Nat. Med.* **14**, 413–420 (2008).
5. Dolan, M.J. *et al. Nat. Immunol.* **8**, 1324–1336 (2007).
6. Townson, J.R., Barcellos, L.F. & Nibbs, R.J. *Eur. J. Immunol.* **32**, 3016–3026 (2002).
7. Locke, D.P. *et al. Am. J. Hum. Genet.* **79**, 275–290 (2006).
8. Khaja, R. *et al. Nat. Genet.* **38**, 1413–1418 (2006).
9. Redon, R. *et al. Nature* **444**, 444–454 (2006).
10. Detels, R. *et al. J. Acquir. Immune Defic. Syndr.* **7**, 1263–1269 (1994).
11. Martin, M.P. *et al. Science* **282**, 1907–1911 (1998).
12. Hendrickson, S.L. *et al. J. Acquir. Immune Defic. Syndr.* **48**, 263–271 (2008).

## Experimental aspects of copy number variant assays at *CCL3L1*

### To the Editor:

Copy number variants (CNVs) are duplicated or deleted segments of the genome that vary in size from a few bases to several kilobases and comprise a substantial proportion of normal genomic variation<sup>1</sup>. The role of population-wide CNVs in disease has only recently come under investigation<sup>2,3</sup>. The gene encoding the chemokine (C-C) motif receptor CCR5, *CCR5*, on chromosome 3p21, has been associated with resistance to HIV-1 infection<sup>2</sup>. One of its ligands, *CCL3L1*, is encoded by a gene that lies in a CNV on chromosome 17q12 (ref. 4), which includes another CCR5 ligand, *CCL4L1* (Supplementary Fig. 1), both of which have been reported to be associated with HIV-1/AIDS susceptibility<sup>2,5,6</sup>. *CCR5* is associated with type 1 diabetes<sup>7</sup>, and hence we hypothesized that *CCL3L1* is also associated with type 1 diabetes.

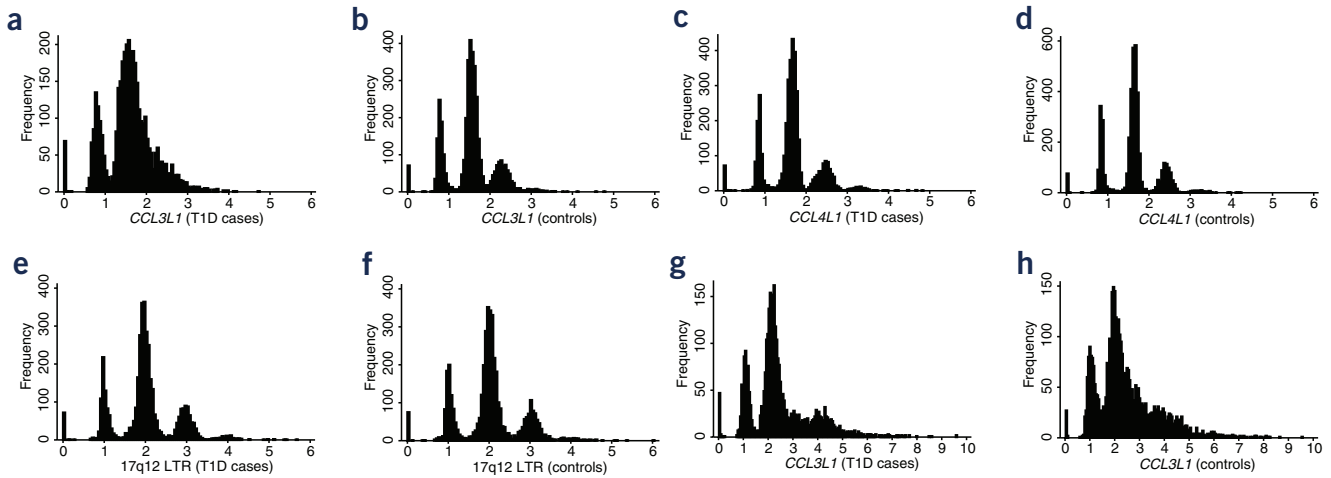
A reliable method for determining copy number at the 17q12 locus is required, as not only has the region been implicated in HIV-1/AIDS susceptibility, but also has been reported to influence disease progression with and without antiretroviral therapy<sup>8,9</sup> and has been suggested as an informative approach to optimizing the design and evaluation of HIV-1 vaccine trials and prevention programs<sup>10</sup>. Quantitative real-time PCR (qPCR) is considered the 'gold standard' method for assessing copy number at individual CNV loci and was employed by the original *CCL3L1* in HIV-1/AIDS study<sup>2</sup> as well as by subsequent studies of this CNV<sup>11</sup>. However, the *CCL3L1* association with HIV-1 has not been independently replicated, with all reported positive associations originating from almost the same case sample set and control sample set<sup>12</sup>. Consequently, there may be experimental biases in the current qPCR *CCL3L1* assay and its scoring.

In type 1 diabetes the median effect of susceptibility loci is below an odds ratio of 1.5 (ref. 13), so thousands of samples are required to test for association, which is also the case for most other complex diseases. Hence, the *CCL3L1* assay has to work efficiently and accurately in thousands of samples. We compared two methods of obtaining *CCL3L1* copy number, the paralogue ratio test (PRT)<sup>14</sup>, and qPCR<sup>2</sup>. PRT and qPCR rely on comparing the signal from the CNV against that from a reference locus and obtaining the ratio. If the reference has two copies, then a ratio of 1:1 denotes two copies of the CNV (assuming equal PCR efficiencies for

both reactions). PRT uses a locus paralogous to the CNV with invariant copy number as the reference<sup>14</sup>. For the *CCL3L1* region, we used three PRT assays with CNVs *CCL3L1* and *CCL4L1* and a long terminal repeat (LTR) located between them and paralogous loci *CCL3* and *CCL4* and a LTR on chromosome 10q22, respectively. We scored these ratios to give integer copy numbers.

We studied DNA samples from 5,771 British individuals with type 1 diabetes and 6,854 geographically matched controls for *CCL3L1* variation using qPCR and PRT assays (Supplementary Methods). Owing to small, but potentially crucial, variations in PCR efficiency, the ratios of assay product formed a distribution around the whole copy number (Fig. 1). With PRT, discrete clusters were distinguishable for the ratios from both *CCL4L1* and the nearby LTR, whereas clusters overlapped for *CCL3L1* (Fig. 1a–f). This could be the result of sequence-specific DNA-bound protein interfering with the PCRs for *CCL3L1*. Differences in the DNA extraction methods for case and control DNA may have left different amounts of DNA-bound protein, resulting in differential cluster quality. In contrast to the LTR, both the *CCL3L1* and *CCL4L1* assay ratios were not centered on integer values but were shifted toward lower values. In controls, for example, one copy of *CCL3L1* was centered on 0.8, two copies were centered on 1.6 and three copies were centered on 2.3 (Fig. 1b). Having examined the distributions of the assay ratios, we assigned integer copy numbers using two methods; one was *k*-means clustering and the other was rounding the PRT data to the nearest integer (Supplementary Methods and Supplementary Discussion). The original HIV/AIDS study used rounding for their qPCR assays<sup>2</sup>.

Except in highly stratified populations, deviation from Hardy-Weinberg equilibrium (HWE) provides a useful indicator of genotyping error. Therefore, we developed a statistical test for HWE with multiallelic CNVs (Supplementary Methods). In controls, *CCL3L1* was in HWE for the *k*-means PRT data but not for the rounded PRT data (Supplementary Table 1), reflecting the inappropriateness of rounding in this case. Both *CCL4L1* and the LTR were in HWE in controls (Table 1 and Supplementary Tables 1 and 2). As the *k*-means clustered data for each assay individually and averaged across assays were in HWE, we tested them for association with type 1 diabetes using a logistic regression



**Figure 1** Histograms of the frequency of the chromosome 17q12 CNV measure obtained using PRT and qPCR assays. Ratios of the CNV (*CCL3L1*, *CCL4L1* and LTR17) versus the reference (paralogous loci *CCL3*, *CCL4* and LTR10 for PRT and the hemoglobin- $\beta$  gene for qPCR) that are used to assign copy numbers, either by rounding the data to the nearest integer or by using *k*-means clustering, are presented. (a,b) Assay ratio (copies) of *CCL3L1* obtained using PRT in 3,860 cases (a) and 4,084 controls (b). (c,d) Assay ratios (copies) of *CCL4L1* obtained using PRT in 4,041 cases (c) and 4,318 controls (d). (e,f) Assay ratios (copies) of the LTR on chr17q12 obtained using PRT in 4,044 cases (e) and in 4,266 controls (f). (g,h) Assay ratios (copies) of *CCL3L1* obtained using qPCR in 3,362 cases (g) and in 3,983 controls (h). T1D, type 1 diabetes.

model, with disease status as the outcome variable and copy number as the independent variable (Supplementary Methods). We found no convincing evidence of association ( $P > 0.05$ ; Table 1, Supplementary Tables 1 and 2 and Supplementary Results). The *CCL3L1* rounded data showed evidence of association with type 1 diabetes ( $P = 8 \times 10^{-11}$ ), but the deviation from HWE suggests that this was false and attributable to genotyping error. We also obtained evidence of association with the rounded data at *CCL4L1* ( $P = 0.0002$ ), owing to an excess of two copies at *CCL4L1*. The majority of the three-copy cluster lay between 2.0 and 2.5 and so was incorrectly scored as two copies when rounded (Fig. 1c,d). Therefore, this association was an artifact of the method of assigning whole copy number (which was rounding) and, consequently, is likely to be false. We concluded that the LTR assay was the most robust of the three assays because the LTR ratios clustered well around integer copy numbers and so could be assigned by rounding or *k*-means, and the assay gave consistent results for the replicated quality control samples. Consequently, we tested the LTR for interaction with the *CCR5Δ32* allele in type 1 diabetes (Supplementary Methods), as combinations of *CCR5-CCL3L1* genotypes have been reported to be associated with HIV-1/AIDS risk and progression<sup>29</sup>. We did not find evidence of a statistical interaction between the associations of *CCR5* and *CCL3L1* with type 1 diabetes ( $P = 0.29$ ).

The distribution of ratios obtained by qPCR for *CCL3L1* in cases and controls was shifted right toward higher than integer copy numbers (Fig. 1g,h). We used *k*-means clustering (Supplementary Methods) and rounding to assign integer copy number. With both methods of assigning copy number, the *CCL3L1* data from qPCR deviated from HWE in controls, owing to a lack of ‘0’ copy numbers, and a rightward shift in the copy number distribution (Supplementary Table 3). There

was evidence for strong associations of *CCL3L1* with type 1 diabetes ( $P = 7 \times 10^{-34}$  (*k*-means) and  $P = 3 \times 10^{-7}$  (rounding)), which, in light of the HWE tests and our PRT results, we regard as artifacts.

For qPCR, we used a standard curve on each plate to standardize the concentration of DNA between CNV and reference reactions within a plate, which introduced plate-to-plate variation. Standard curves are not required for PRT, which has reference and CNV reactions in the same well. As cases and controls were dispensed onto separate plates, the statistical test for association may have actually detected plate-to-plate variation and not type 1 diabetes association. Because we had so many plates (105), we were able to estimate the qPCR plate-to-plate variation (6% in controls and 2% in cases) to correct the association test. We reduced the apparent evidence of type 1 diabetes association,  $P = 7 \times 10^{-9}$  (*k*-means; Table

**Table 1** Copy number of the 17q12 CNV.

Copy number	LTR (PRT)		<i>CCL3L1</i> (qPCR)	
	Obs. (exp.)	Obs. (exp.)	Obs. (exp.)	Obs. (exp.)
	Cases	Controls	Cases	Controls
0	75 (62.2)	78 (73.9)	53 (56.9)	29 (68.3)
1	726 (752.3)	829 (837.4)	598 (590.9)	750 (668.6)
2	2408 (2393.3)	2510 (2505.0)	1610 (1610.5)	1729 (1771.8)
3	732 (733.7)	756 (757.0)	467 (461.5)	746 (739.8)
4	93 (92.3)	83 (82.6)	399 (400.3)	453 (451.5)
5	9 (9.4)	9 (9.1)	158 (169.3)	173 (175.8)
6	1 (0.8)	1 (0.9)	45 (51.8)	66 (80.9)
7			25 (15.9)	23 (18.8)
8			6 (4.2)	12 (6.0)
9			1 (0.7)	1 (1.2)
10			0 (0.0)	1 (0.3)
$P_{HWE}$	0.1562	0.8242	0.1454	$1 \times 10^{-8}$
$P_{T1D}$	0.6946		0.0002	

Copy numbers were obtained using the LTR PRT assay in 4,044 cases and 4,266 controls, and from the qPCR assay in 3,362 cases and 3,983 controls (2,479 cases and 2,378 controls were successfully genotyped with both assays). Obs., observed counts; exp., expected counts;  $P_{HWE}$ , Hardy-Weinberg equilibrium *P* value;  $P_{T1D}$ , *P* value for association with type 1 diabetes.

1) and  $P = 0.00019$  (rounding; **Supplementary Table 3**). The remaining evidence of association may be attributable to shifts in copy number distribution, caused by unpredictable interactions between the assay and the differential quality and composition of DNA from several sources, leading to HWE deviation in the controls but not the cases<sup>15</sup> (**Supplementary Discussion**). These results not only show the importance of testing for HWE but also of allowing for plate effects in the analysis. We recommend arraying case and control samples onto the same plates.

5,121 samples were common to the PRT and qPCR experiments. Using the LTR as the most accurately scored measure of *CCL3L1* for the PRT method and the *k*-means clustered data for the qPCR assay, we found that 64% were consistent between the two methods, with 25% having one additional copy of *CCL3L1* with qPCR as compared to with PRT. Nine percent of the data had between two and five additional copies of *CCL3L1* obtained using qPCR compared to PRT. Just 2% of the data had one or two copies more of *CCL3L1* as measured by PRT than by qPCR. The qPCR assay also showed a general trend toward higher copy numbers compared to the other two PRT assays; a shift that may be an artifact of the qPCR primers also binding a *CCL3L1* pseudogene, *CCL3L2*, as well as *CCL3L1* (refs. 2,4,14) (**Supplementary Fig. 1** and **Supplementary Discussion**).

To our knowledge, this is the first report where *CCL3L1* copy number has been estimated in such a large sample set using both qPCR and PRT. The PRT LTR assay copy number showed good clustering, with no difference in distribution between type 1 diabetes cases and controls, suggesting that the primers designed for this locus are highly specific and robust to variations in source of DNA. PRT also avoids potential error in scoring zero copy numbers. We recommend rounding the data if the assay ratios cluster distinctly around integers and using *k*-means clustering otherwise. qPCR can be used to assay *CCL3L1* in large well-powered sample sets if appropriate quality control measures are implemented. The distribution of copy number and its dependence on DNA source should be examined statistically, and an appropriate method of assigning copy number should be adopted for each DNA source. Tests for deviation from HWE must be performed, and any detected deviations should be resolved. If they cannot be resolved, an assay such as PRT should be used instead. In small sample sets (for example,  $n \leq 500$ ), even in the absence of deviations from HWE, any associations should be treated with some skepticism, owing to the limited power to detect deviations from HWE (**Supplementary Results**).

Finally, we note that there is marked variation in *CCL3L1* copy number according to ethnic group, which others have reported<sup>2</sup> (**Supplementary Table 4**). We genotyped 95 African Yoruban samples in duplicate using the LTR PRT assay. Our data was highly reproducible (correlation coefficient  $> 0.99$ ; **Supplementary Fig. 2**). The CNVs were between 2.0 and 8.0 (**Supplementary Table 4**), with a mean copy number of 4.3. Hence, any inadvertent admixture, as seems possible upon detailed evaluation of the original *CCL3L1* in HIV-1/AIDS study in which European Americans had different *CCL3L1* copy numbers than Hispanic Americans<sup>2</sup>, combined with the copy number distribution shifts interacting with DNA source and the error-prone copy number scoring (that is, rounding) described here, could lead to apparently highly significant disease associations.

We collected all DNA samples with approval from the Cambridgeshire 2 Research Ethics Committee, and we obtained written consent from all individuals or from the parents of individuals who were too young to consent.

*Note: Supplementary information is available on the Nature Medicine website.*

Sarah F Field<sup>1,5</sup>, Joanna M M Howson<sup>1,5</sup>, Lisa M Maier<sup>2,3,5</sup>, Susan Walker<sup>4</sup>, Neil M Walker<sup>1</sup>, Deborah J Smyth<sup>1</sup>, John A L Armour<sup>4</sup>, David G Clayton<sup>1</sup> & John A Todd<sup>1</sup>

<sup>1</sup>Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, Department of Medical Genetics, University of Cambridge, Cambridge, UK. <sup>2</sup>Division of Molecular Immunology, Center for Neurologic Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. <sup>3</sup>Program in Medical and Population Genetics, Broad Institute, Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA. <sup>4</sup>Institute of Genetics and School of Biology, University of Nottingham, Queen's Medical Centre, Nottingham, UK. <sup>5</sup>These authors contributed equally to this work.  
e-mail: John.Todd@cimr.cam.ac.uk

#### ACKNOWLEDGMENTS

This work was funded by the Juvenile Diabetes Research Foundation International, the Wellcome Trust and the UK National Institute for Health Research Cambridge Biomedical Research Center. The Cambridge Institute for Medical Research is in receipt of a Wellcome Trust Strategic Award (079895). We are grateful for the participation of the individuals with type 1 diabetes and the control individuals. We would like to thank the UK Medical Research Council and Wellcome Trust for funding the collection of DNA for the British 1958 Birth Cohort. DNA control samples were prepared and provided by S. Ring, R. Jones, M. Pembrey and W. McArdle (Avon Longitudinal Study of Pregnancy and Childhood Laboratory, Department of Social Medicine, University of Bristol), D. Strachan (Division of Community Health Sciences, St. George's, University of London) and P. Burton (Department of Genetics and Health Sciences (University of Leicester)). Type 1 diabetes case DNA samples were prepared by K. Bourget, S. Duley, S. Hawkins, G. Coleman, M. Maisuria, S. Hood, E. King, T. Mistry, A. Simpson, S. Wood, S. Clayton, F. Wright and H. Stevens (Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge). Double scoring was performed by M. Hardy (Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge).

1. McCarroll, S.A. *et al. Nat. Genet.* **40**, 1166–1174 (2008).
2. Gonzalez, E. *et al. Science* **307**, 1434–1440 (2005).
3. McCarroll, S.A. *Hum. Mol. Genet.* **17**, R135–R142 (2008).
4. Modi, W.S. *Genomics* **83**, 735–738 (2004).
5. Zimmerman, P.A. *et al. Mol. Med.* **3**, 23–36 (1997).
6. Colobran, R. *et al. J. Immunol.* **174**, 5655–5664 (2005).
7. Smyth, D.J. *et al. N. Engl. J. Med.* **359**, 2767–2777 (2008).
8. Dolan, M.J. *et al. Nat. Immunol.* **8**, 1324–1336 (2007).
9. Ahuja, S.K. *et al. Nat. Med.* **14**, 413–420 (2008).
10. Kulkarni, H. *et al. PLoS One* **3**, e3671 (2008).
11. Mamtani, M. *et al. Ann. Rheum. Dis.* **67**, 1076–1083 (2008).
12. Shao, W. *et al. Genes Immun.* **8**, 224–231 (2007).
13. Todd, J.A. *et al. Nat. Genet.* **39**, 857–864 (2007).
14. Walker, S., Janyakhantikul, S. & Armour, J.A. *Genomics* **93**, 98–103 (2009).
15. Clayton, D.G. *et al. Nat. Genet.* **37**, 1243–1246 (2005).

#### He *et al.* reply:

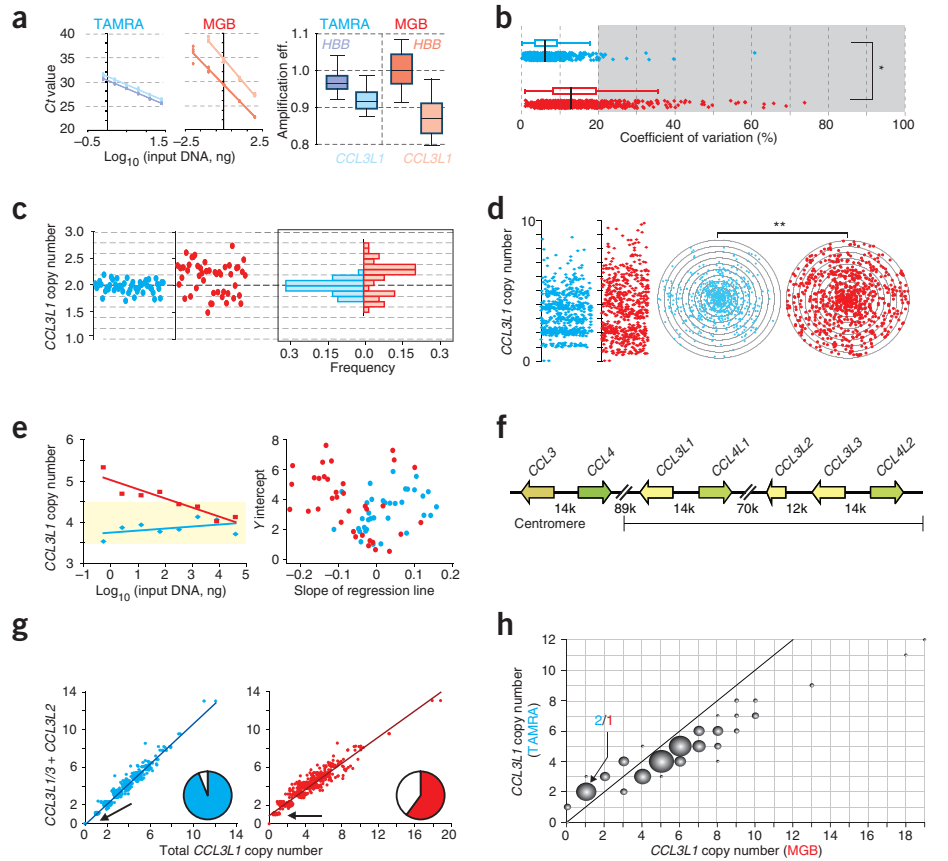
We read with interest the three correspondences<sup>1–3</sup> that are directly or indirectly related to our previous publication in *Nature Medicine*<sup>4</sup>. Because of methodological concerns with their own C-C chemokine ligand-3-like-1 (*CCL3L1*) gene quantitative PCR (qPCR) assay, Urban *et al.*<sup>1</sup> assume that similar limitations apply to the assay that we and others have used and, by extension, infer that all of the several reported associations for *CCL3L1* copy number with HIV/AIDS susceptibility (**Supplementary Table 1**) are spurious. We tested the validity of this assumption. Notwithstanding their

statement that they “assayed for *CCL3L1* copy number using a previously described method” (described in ref. 5 here), Urban *et al.*<sup>1</sup> made two major modifications to our qPCR assay, which we show adversely affect assay performance. Although they used the same probe sequences, instead of using tetramethylrhodamine (TAMRA) as the quencher, they used minor groove binder (MGB), a change that raises the melting temperature of the probe 8 °C higher than the optimal temperature (**Supplementary Table 2**). Urban *et al.*<sup>1</sup> also raise the concern that our assay is confounded, because they observed that *CCL3L1* gene copy number (GCN) differed with the



**Figure 1** Comparison of the performance of the TAMRA assay<sup>5</sup> (color-coded blue) and MGB assay<sup>1</sup> (color-coded red) using identical human DNA samples in concurrently run experiments as described in the **Supplementary Methods**.

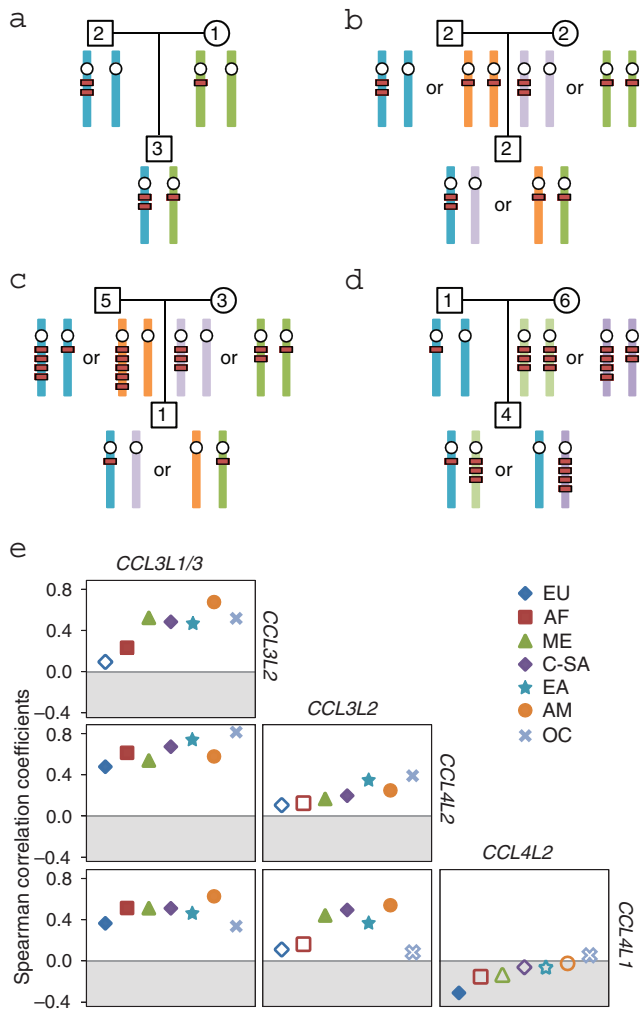
**(a)** Efficiency of assay. Left, representative standard curves of serial dilutions for the target (*CCL3L1*; light color) and normalizer (*HBB*; dark color) genes using the TAMRA assay and the MGB assay. Right, box plots depicting amplification efficiencies (eff.) (TAMRA  $n = 27$ ; MGB  $n = 26$  independent experiments). **(b)** Intra-assay reproducibility, as determined using triplicates for each of the 628 DNA samples. Box plot representation of the coefficient of variation is also depicted.  $*P = 5.4 \times 10^{-58}$  (paired  $t$  test). Gray shading indicates coefficient of variations exceeding 20%. The intra-assay repeatability of *CCL3L1* GCN from 184 DNA samples was also higher for the TAMRA assay compared to the MGB assay (**Supplementary Table 4**). **(c)** Accuracy of assay to estimate the authenticated two *CCL3L1* copies in the A431 cell line. The assay was conducted six separate times with eight serial dilutions each, yielding 48 *CCL3L1* GCN estimates (scatter plots on the left). Differences in the accuracies were examined using a bihistogram, shown on the right. The mean of the *CCL3L1* GCN obtained by TAMRA assay (1.98) and MGB assay (2.15) differed significantly ( $P = 8.8 \times 10^{-5}$ ), as did the standard deviations (0.11 and 0.30 for TAMRA assay and MGB assay, respectively;  $P = 3.8 \times 10^{-10}$ ). **(d)** Precision of assay, defined as the ability of the assay to provide an estimated GCN close to an integer value. Left, scatter plots for the unrounded GCN estimates (truncated to ten copies for each assay). Right, clustering around integers, as assessed by estimating the Euclidean distance of the unrounded *CCL3L1* GCN estimate from the nearest integer. The center represents a distance of zero, whereas the outermost circle represents a distance of 0.5. The concentric circles represent distances in multiples of 0.05.  $**P = 3.9 \times 10^{-31}$ . **(e)** Sensitivity of the assay to input DNA, as examined in 35 samples, each with eight serial dilutions (100 to 0.78 ng). Left, example from a DNA sample with four copies. A slope of zero indicates no influence of the amount of input DNA on the final estimate of the copy number. In this particular example, the slope was significantly different from zero for the MGB assay (slope,  $-0.222$ ;  $P = 0.001$ ) but not TAMRA assay (slope,  $0.048$ ;  $P = 0.287$ ). Yellow shading encompasses unrounded *CCL3L1* GCN estimates (3.5–4.5) that would be categorized as four copies. Right, results of the regression analyses for 35 samples. The slope ( $x$  axis) and  $y$  intercept ( $y$  axis) derived from the regression analyses of eight serial dilutions for each of the 35 samples (dots) are plotted. The differences in the slopes (mean, 95% confidence interval (CI)) for the TAMRA assay ( $0.02$ ;  $-0.01$  to  $0.05$ ) and the MGB assay ( $-0.07$ ;  $-0.10$  to  $-0.05$ ) were statistically different ( $P = 5.7 \times 10^{-6}$ ). **(f)** Schematic of the genomic architecture of the chromosome 17q12 region, depicting the *CCL3L* and *CCL4L* genes<sup>14</sup>. **(g)** Internal validation of the assays, reflected by the concordance between unrounded total *CCL3L1* GCN (determined by TAMRA or MGB assay) and sum of the unrounded copy numbers of *CCL3L1* and *CCL3L3* (*CCL3L1/3*) plus *CCL3L2*, determined by two separate assays (the assay cannot distinguish between *CCL3L1* and *CCL3L3* (refs. 5 and 7);  $n = 628$  samples). Arrows show that the intercepts for the regression lines for the TAMRA assay and MGB assay differ ( $R^2 = 0.9561$  ( $y = 1.074x - 0.104$ ) for TAMRA assay;  $R^2 = 0.9055$  ( $y = 0.6897x + 0.9075$ ) for MGB assay). Insets, blue and red pie charts depict concordance between rounded total *CCL3L1* GCN and sum of the rounded copy numbers of *CCL3L1/3* and *CCL3L2* in 184 samples (**Supplementary Table 5**). **(h)** Bubble plot showing the disagreement between rounded *CCL3L1* GCN estimates made by the TAMRA ( $y$  axis) and MGB ( $x$  axis) assays. Each bubble is centered over the intersection of the specific discrepant GCN values, and the size of the bubble is proportional to the number of discrepant samples at each intersection. The angled line indicates rounded *CCL3L1* GCNs that were identical for both assays, and, for ease of visualization, these samples are not depicted with bubbles (351 of 628 samples (56%)). The MGB assay underestimated in 62 (10%) and overestimated in 215 (34%) samples the *CCL3L1* GCN compared to the *CCL3L1* GCN from the TAMRA assay (**Supplementary Table 6**). Bubble 2/1, two copies estimated by TAMRA assay, one copy by MGB assay.



quality and the amount of input DNA. Our published protocol specifically recommended a fixed amount of high-quality input DNA between 2 and 10 ng<sup>5</sup>, and, consequently, we used qPCR standard curves comprising twofold dilutions from 25 to 0.78 ng. By contrast, Urban *et al.*<sup>1</sup> used variable amounts of input DNA (up to 100 ng) and constructed standard curves of tenfold dilutions from 100 ng to 1 pg.

The TAMRA-based assay that was initially used to estimate *CCL3L1* GCN<sup>6</sup> was, as reported previously<sup>5</sup>, refined by us until its performance met rigorous quality control measures similar to those shown in **Figure 1**. Here we concurrently evaluated the performances of the MGB and TAMRA-based assays. The standard curves for the target (*CCL3L1*) and normalizer (*HBB*; encoding human  $\beta$ -globin B) genes were parallel with

the TAMRA assay, whereas this was not the case for the MGB assay (**Fig. 1a**). Consequently, the inferred amplification efficiencies for *CCL3L1* and *HBB* were more divergent for the MGB assay than for the TAMRA assay (**Fig. 1a**). Compared with the MGB assay, the TAMRA assay had better intra-assay reproducibility (**Fig. 1b**). The unrounded *CCL3L1* GCN estimates in a cell line authenticated to have two copies (**Fig. 1c**) and in 628 human DNA samples (**Fig. 1d**) were significantly closer to the nearest integer with the TAMRA assay than with the MGB assay. In agreement with the findings of Urban *et al.*<sup>1</sup>, the *CCL3L1* GCN with the MGB assay differs with varying input DNA (**Fig. 1e**). However, this sensitivity to input DNA is not a characteristic of the TAMRA assay that we used (**Fig. 1e**).



**Figure 2** Chromosome phase of *CCL3L1* GCN in parent-child trios and correlations among *CCL3L* and *CCL4L* genes. (a–d) Predicted *CCL3L1*-GCN chromosome phase in parent-child trios. Representative examples from 655 parent-child trios showing *CCL3L1*-GCN phase resolved (a) and unresolved (b–d) pedigrees. The chromosomes are color coded to denote possible inheritance patterns. Numbers inside the symbols are the total *CCL3L1* GCN, and the possible GCN phase is represented by the boxes along the chromosome. Centromeres are depicted as white circles. (e) Matrix of the pair-wise Spearman correlation coefficients ( $\rho$  values) among the copy numbers of *CCL3L1/3*, *CCL3L2*, *CCL4L1* and *CCL4L2* genes. Symbols reflect HGDP-CEPH samples (EU, Europe ( $n = 161$ ); AF, Africa ( $n = 127$ ); ME, Middle East ( $n = 178$ ); C-SA, Central-South Asia ( $n = 196$ ); EA, East Asia ( $n = 251$ ); AM, America ( $n = 108$ ); and OC, Oceania ( $n = 39$ )). Open and filled symbols reflect  $P$  values that were nonsignificant and significant (Supplementary Table 7), respectively, for the pair-wise comparisons of the correlations.

We previously showed the internal validity of the *CCL3L1* GCN obtained by the TAMRA assay<sup>7</sup>. *CCL3L1* GCN reflects the sum of the GCNs of three separate *CCL3L* genes (*CCL3L1*, *CCL3L2* and *CCL3L3*; Fig. 1f). We quantified the GCNs of these three *CCL3Ls* with two separate assays, one specific for *CCL3L1* and *CCL3L3* GCN and the other for *CCL3L2* GCN (Fig. 1f)<sup>7</sup>. Thus, a ‘total’ *CCL3L1* GCN determination has internal validity if it approximates the sum of the GCNs of the individual *CCL3Ls*<sup>7</sup>. This, we found, was much more consistently the case with the TAMRA assay than with the MGB assay (Fig. 1g). Consequently, the concordance between the rounded (nearest integer) *CCL3L1* and *CCL3L3* GCN plus *CCL3L2* GCN and ‘total’ *CCL3L1* GCN was 94% with the TAMRA assay,

whereas it was 60% with the MGB assay (Fig. 1g). In keeping with these divergent results, we observed discrepancies between the TAMRA and MGB assays in the rounded *CCL3L1* GCN in 44% of 628 samples. The impact of these misclassifications was amplified by the fact that, relative to the TAMRA assay, the MGB assay underestimated GCN at lower values but overestimated it at higher GCN values (Fig. 1h). Given the performance characteristics of the two assays (Fig. 1a–g), it is reasonable to conclude that the true GCN corresponds better to the TAMRA assay value. Notably, among the 177 samples classified by TAMRA as having two copies, the MGB assay underestimated 37 samples as bearing one copy (Fig. 1h), an error that may obscure the association of low copy numbers with HIV/AIDS susceptibility. For example, we previously found that a low copy number of *CCL3L1* (assayed by TAMRA probe) associated with increased HIV/AIDS susceptibility in 298 Ukrainian children exposed perinatally to HIV-1<sup>7</sup>; affirming this, we observed similar associations for *CCL4L1* GCN (determined by a separate assay)<sup>7</sup>, a predictable association because, in this study population, the copy numbers of *CCL3L1* and *CCL4L1* are highly correlated<sup>7</sup>. In contrast, when we repeated the genotyping of *CCL3L1* GCN in this cohort with the MGB-based assay, we found that *CCL3L1* GCN estimated by the MGB assay failed to identify the associations of *CCL3L1* GCN with the risk of acquiring HIV infection (odds ratio, 0.94; 95% confidence interval, 0.56–1.56;  $P = 0.80$ ) and rate of progression to AIDS (relative hazard, 1.41; 95% confidence interval, 0.79–2.51;  $P = 0.24$ ). Thus, we surmise that the use of the MGB assay and variable input DNA together may have compromised the results of the association studies by Urban *et al.*<sup>1</sup>

Bhattacharya *et al.*<sup>2</sup> also modified the qPCR assay (Supplementary Table 2). The predicted melting temperature for their Quencher Series 7 (QSY7)-based probe is 8 °C lower than the optimal temperature. Although we have had insufficient time since receiving this correspondence to make direct comparisons between the QSY7 assay and the TAMRA assay, conceivably this modification may affect assay performance. Moreover, Bhattacharya *et al.*<sup>2</sup> have not adequately accounted for genetic-epidemiological factors pertinent to the full interpretation of their results. First, in subjects followed from the early stages of HIV infection, we found that, because a high *CCL3L1* GCN associated with a slower rate of progression to AIDS and death, at a population level there was a gradual enrichment over time of HIV-positive subjects with a protective high *CCL3L1* GCN<sup>5</sup>. Consequently, ~7 years after seroconversion, the prevalence of *CCL3L1* GCN in surviving HIV-positive individuals approaches that of HIV-negative individuals<sup>5</sup>. Hence, because Bhattacharya *et al.*<sup>2</sup> studied mostly seroprevalent subjects (~78%), it is predictable that they found similar *CCL3L1* GCN distributions in HIV-infected and uninfected subjects. Second, during the study period (12 years) there were only 53 AIDS events, representing 13% of their subjects ( $n = 396$ ). By contrast, during a comparable time period, AIDS developed in ~40% of the 1,132 adult subjects we studied<sup>5</sup>. Given the limited AIDS events and small sample size, it is not surprising that strong associations between *CCL3L1* GCN and AIDS were not detected. Nevertheless, their data confirm our results showing that long-term nonprogressors are enriched for a high *CCL3L1* GCN<sup>8</sup>. Third, Bhattacharya *et al.*<sup>2</sup> failed to detect an association between *CCL3L1* GCN or conjoint genotypes of *CCL3L1* GCN and the *CCR5Δ32* allele with highly active antiretroviral therapy (HAART) responses. They suggested that the observed lack of association of *CCL3L1* GCN with recovery of CD4<sup>+</sup> T cell counts during HAART was robust, because they found that *CCR5Δ32* heterozygosity also did not associate with beneficial CD4<sup>+</sup> T cell count responses in their cohort<sup>2</sup>. However, contrary to their results, several studies have shown that *CCR5Δ32* heterozygosity is associated with improved CD4<sup>+</sup> T cell and viral load responses during HAART<sup>9–13</sup>, including work from the Multicenter AIDS Cohort Study cohort, from which their study subjects were derived<sup>2</sup>. Thus, their failure to find an

association between *CCR5Δ32* heterozygosity and HAART responses does not confer or imply validity for the lack of an association for *CCL3L1* GCN with CD4<sup>+</sup> T cell count responses, as both of these findings reported by Bhattacharya *et al.*<sup>2</sup> may represent false negatives.

Notably, Fields *et al.*<sup>3</sup>, on the basis of their results using a paralogous ratio test (PRT) and qPCR assay for *CCL3L1* GCN conducted in subjects with diabetes, suggest that our observed associations of *CCL3L1* GCN with HIV/AIDS are suspect. However, their PRT assay quantifies two *CCL3L* genes (*CCL3L1* and *CCL3L3*), whereas the qPCR we used for association studies quantifies three<sup>5,7</sup> (Fig. 1f). This precludes any direct comparison of the *CCL3L1* GCN estimates obtained by PRT versus qPCR and limits inferences of association studies because it does not account for the possible contributions to disease susceptibility of *CCL3L2*, which, contrary to previous assumptions, does not seem to be a pseudogene<sup>7</sup>.

Because Fields *et al.*<sup>3</sup> found that the *CCL3L1* GCN departed from Hardy-Weinberg equilibrium, they infer that our qPCR assay does not provide robust GCN estimates. This inference based on Hardy-Weinberg estimates oversimplifies the analysis of a highly complex genomic region. Their Hardy-Weinberg estimates would be valid if the copy number variations (CNVs) of the individual *CCL3L* genes were both duplicated and distributed randomly on each chromosome, permitting accurate computation of copies per chromosome. However, recent studies<sup>14,15</sup> and the following four data sets suggest otherwise. First, we derived *CCL3L1* GCN data from 655 parent-child trios and found that it is very difficult to predict the chromosome phase of *CCL3L1* GCN, even in individuals with two copies but especially in those with >2 copies (Fig. 2a–d). Second, consistent with this observation, using the expectation maximization algorithm described by Fields *et al.*<sup>3</sup>, we found that Hardy-Weinberg estimates in 991 normal donors departed from equilibrium only when they were computed for a *CCL3L1* GCN of >2 but not ≤2 (Supplementary Note). Third, among 655 parent-child trios, we identified 62 trios for whom the *CCL3L1* GCN phase of each family member could be resolved. However, the expectation maximization algorithm misclassified the *CCL3L1* GCN phase in 13% of these 186 subjects. For example, 20 individuals with two gene copies of *CCL3L1* on one chromosome and zero copies on the other chromosome were misclassified as having one gene copy of *CCL3L1* on each chromosome (Supplementary Table 3).

Finally, underscoring that the *CCL3L-CCL4L* region is unlikely to be a single duplicon, there are fewer *CCL4L* than *CCL3L* copies<sup>6,7,14</sup>, and correlations of the GCN among *CCL3Ls* and *CCL4Ls* vary significantly among populations represented in the Human Genome Diversity Cell Line Panel (HGDP)-CEPH samples (Fig. 2e). In some populations, there is an inverse correlation between *CCL4L1* and *CCL4L2* GCNs (Fig. 2e). Consistent with this unbalanced distribution of *CCL3L* and *CCL4L* genes, high-resolution CNV data also show that the *CCL3L-CCL4L* region shows extensive architectural complexity, with smaller CNVs embedded within larger ones and with interindividual variation in breakpoints<sup>15</sup>. This degree of complexity is also highlighted by recent sequence data showing that the most extreme copy number variation corresponds to genes that are embedded within segmental duplications<sup>16</sup>, such as the *CCL3L-CCL4L* genes (Fig. 1f)<sup>7,17</sup>. Thus, accounting for this genomic complexity, including *CCL3L2* GCN, is crucial for full interpretation of association studies, because the combinatorial content of *CCL3L* and *CCL4L* GCNs influences HIV/AIDS susceptibility<sup>7,14</sup>. On a technical note, there was a very high correlation between the rounded versus the *k*-mean clustered *CCL3L1* GCN estimates (Supplementary Note), making it unlikely that the method used to discretize GCN confounded our association studies.

Contrary to the assertions in these three correspondences<sup>1–3</sup>, the association of *CCL3L1* GCN with HIV/AIDS susceptibility has been confirmed by us and others in multiple cohorts and also in nonhuman primates<sup>17,18</sup> (Supplementary Table 1). Moreover, there is a highly plau-

sible biological and experimental basis for these associations: *CCL3L1* is the most potent HIV-suppressive chemokine, and a low *CCL3L1* GCN associates with lower chemokine expression, reduced cell-mediated immunity, lower HIV-specific Gag responses and higher viral loads (Supplementary Table 1). Furthermore, concordant associations of *CCL3L1-CCR5* genotypes with immune reconstitution were detected among subjects who received HAART at three stages of HIV disease: acute, early and chronic infection<sup>4</sup>. Notably, among subjects who received HAART during chronic infection, consistent associations for *CCL3L1-CCR5* genotypes were observed for two phenotypic endpoints, namely, recovery of CD4<sup>+</sup> T cell counts<sup>4</sup> and CD4<sup>+</sup> T cell function<sup>8</sup>. We surmise that methodological and epidemiological confounders and failure to account for the genomic complexity of the *CCL3L-CCL4L* locus underlie the lack of associations reported by these correspondences<sup>1–3</sup>.

The human DNA samples used were in accordance with the Institutional Review Boards of University of Texas Health Science Center and University of California at San Diego.

Note: Supplementary information is available on the Nature Medicine website.

Weijing He<sup>1,2</sup>, Hemant Kulkarni<sup>1,2</sup>, John Castiblanco<sup>1,2</sup>, Chisato Shimizu<sup>3</sup>, Una Ahuyen<sup>1,2</sup>, Robert Maldonado<sup>1,2</sup>, Andrew Carrillo<sup>1,2</sup>, Madeline Griffin<sup>1,2</sup>, Amanda Lipsitt<sup>1,2</sup>, Lisa Beachy<sup>1,2</sup>, Ludmila Shostakovich-Koretskaya<sup>4</sup>, Andrea Mangano<sup>5</sup>, Luisa Sen<sup>5</sup>, Robert J B Nibbs<sup>6</sup>, Caroline T Tiemessen<sup>7</sup>, Hector Bolivar<sup>8</sup>, Michael J Bamshad<sup>9</sup>, Robert A Clark<sup>1,2</sup>, Jane C Burns<sup>3</sup>, Matthew J Dolan<sup>10,11</sup> & Sunil K Ahuja<sup>1,2</sup>

<sup>1</sup>Veterans Administration Research Center for AIDS and HIV-1 Infection, South Texas Veterans Health Care System, San Antonio, Texas, USA.

<sup>2</sup>Department of Medicine, University of Texas Health Science Center, San Antonio, Texas, USA.

<sup>3</sup>Department of Pediatrics, University of California—San Diego School of Medicine, La Jolla, California, USA.

<sup>4</sup>Department of General Pediatrics and Pediatric Infectious Diseases, Dnepropetrovsk State Medical Academy, Dnepropetrovsk, Ukraine.

<sup>5</sup>Laboratorio de Biología Celular y Retrovirus, Hospital de Pediatría 'J.P. Garrahan', Buenos Aires, Argentina.

<sup>6</sup>Glasgow Biomedical Research Centre, University of Glasgow, Glasgow, Scotland, UK.

<sup>7</sup>AIDS Virus Research Unit, National Institute for Communicable Diseases/National Health Laboratory Services and University of the Witwatersrand, Johannesburg, South Africa.

<sup>8</sup>AIDS Clinical Research Unit, University of Miami, Miller School of Medicine, Miami, Florida, USA.

<sup>9</sup>Departments of Pediatrics and Genome Sciences, University of Washington and Seattle Children's Hospital, Seattle, Washington, USA.

<sup>10</sup>Henry M. Jackson Foundation, Wilford Hall United States Air Force Medical Center, Lackland Air Force Base, Texas, USA.

<sup>11</sup>Infectious Disease Service, San Antonio Military Medical Center, Brooke Army Medical Center, Fort Sam Houston, Texas, USA. e-mail: ahujas@uthscsa.edu

- Urban, T.J. *et al. Nat. Med.* **15**, 1110–1112 (2009).
- Bhattacharya, T. *et al. Nat. Med.* **15**, 1112–1115 (2009).
- Field, S.F. *et al. Nat. Med.* **15**, 1115–1117 (2009).
- Ahuja, S.K. *et al. Nat. Med.* **14**, 413–420 (2008).
- Gonzalez, E. *et al. Science* **307**, 1434–1440 (2005).
- Townson, J.R., Barcellos, L.F. & Nibbs, R.J. *Eur. J. Immunol.* **32**, 3016–3026 (2002).
- Shostakovich-Koretskaya, L. *et al. AIDS* **23**, 679–688 (2009).
- Dolan, M.J. *et al. Nat. Immunol.* **8**, 1324–1336 (2007).
- Guérin, S. *et al. AIDS* **14**, 2788–2790 (2000).
- Yamashita, T.E. *et al. AIDS* **15**, 735–746 (2001).
- Valdez, H., Purvis, S.F., Lederman, M.M., Fillingame, M. & Zimmerman, P.A. *J. Am. Med. Assoc.* **282**, 734 (1999).
- Hendrickson, S.L. *et al. J. Acquir. Immune Defic. Syndr.* **48**, 263–271 (2008).
- Kasten, S. *et al. Eur. J. Med. Res.* **5**, 323–328 (2000).
- Colobran, R. *et al. Genes Immun.* **9**, 279–288 (2008).
- Perry, G.H. *et al. Am. J. Hum. Genet.* **82**, 685–695 (2008).
- Alkan, C. *et al. Nat. Genet.* advance online publication, doi:10.1038/ng.437 (30 August 2009).
- Gornalusse, G. *et al. PLoS Genet.* **5**, e1000359 (2009).
- Degenhardt, J.D. *et al. PLoS Genet.* **5**, e1000346 (2009).