Annals of human genetics

No evidence for association of β-defensin genomic copy number with HIV susceptibility, HIV load during clinical latency, or progression to AIDS

Journal:	*Annals of Human Genetics*
Manuscript ID	AHG-MS-16-0165.R1
Manuscript Type:	Regular manuscript
Date Submitted by the Author:	14-Nov-2016
Complete List of Authors:	Abujaber, Razan; University of Leicester, Department of Genetics Shea, Patrick; Columbia University, Institute for Genomic Medicine McLaren, Paul; Public Health Agency of Canada, National HIV and Retrovirology Laboratory; University of Manitoba, Department of Medical Microbiology and Infectious Diseases Lakhi, Shabir; International Aids Vaccine Initiative; Zambia-Emory HIV Research Project Gilmour, Jill; International Aids Vaccine Initiative; Imperial College London, IAVI Human Immunology Laboratory Allen, Susan; Emory University, Rollins School of Public Health Fellay, Jacques; Ecole Polytechnique Federale de Lausanne, School of Life Sciences Hollox, Edward; University of Leicester, Genetics
Keywords:	β-defensin, copy number variation, CNV, HIV-1, AIDS

No evidence for association of β-defensin genomic copy number with HIV susceptibility, HIV load during clinical latency, or progression to AIDS

Razan Abujaber (1), Patrick Shea (2), Paul J McLaren (3,4), Shabir Lakhi (5,6), Jill Gilmour (5,7), Susan Allen (5,8), Jacques Fellay (9), Edward J Hollox (1), IAVI Africa HIV Prevention Partnership, Swiss HIV Cohort Study

1. Department of Genetics, University of Leicester, Leicester, UK
2. Institute for Genomic Medicine, Columbia University, New York, New York, USA
3. National HIV and Retrovirology Laboratory, Public Health Agency of Canada, Winnipeg, Canada
4. Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, Canada.
5. International AIDS Vaccine Initiative (IAVI), New York, New York, USA
6. Zambia-Emory HIV Research Project, Lusaka and Copperbelt, Zambia.
7. IAVI Human Immunology Laboratory, Imperial College, London, United Kingdom.
8 School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
9. Emory University, Atlanta, Georgia, United States of America.

Corresponding author

Dr Ed Hollox,

Department of Genetics, Adrian Building

University of Leicester

Leicester LE1 7RH, UK

Ejh33@le.ac.uk

+44 116 252 3407

Keywords

β-defensin, copy number variation, CNV, HIV-1, AIDS

Abstract

Common single nucleotide variation in the host accounts for 25% of the variability in the plasma levels of HIV during the clinical latency stage (viral load setpoint). However, the role of rare variants and copy number variants remains relatively unexplored. Previous work has suggested copy number variation of a cluster of β-defensin genes affects HIV load in treatment-naïve sub-Saharan Africans and rate of response to anti-retroviral treatment. Here we analyse a total of 1827 individuals from two cohorts of HIV-infected individuals from Europe and sub-Saharan Africa to investigate the role of β-defensin copy number variation on HIV load at setpoint. We find no evidence ~~of~~ for association of copy number with viral load. We also compare distribution of β-defensin copy number between European cases and controls and find no differences, arguing against a role of β-defensin copy number in HIV acquisition. Taken together, our data argues against an effect of copy number variation of the β-defensin region in the spontaneous control of HIV infection.

Introduction

Rates of HIV acquisition and progression, and levels of viral control during the clinical latency period, show differences between individuals, which are in part due to genetic variation (Shea *et al.*, 2013). The role of gene copy number variation, where the number of copies of the same gene differs between individuals, in affecting clinical parameters of HIV infection is of interest (Hollox & Hoh, 2014). In particular, sequence and copy number variation of the killer immunoglobulin receptor family (KIR) has been shown to be important in control of the progression of HIV (Pelak *et al.*, 2011), particularly in the context of variation at its ligand HLA-B (Bashirova *et al.*, 2011). The role of *CCL3L1* copy number in HIV infection and progression has been much debated (Cantsilieris & White, 2013), and most studies that used robust approaches to measure copy number failed to find any association with likelihood of infection or viral load (Hollox & Hoh, 2014, Aklillu *et al.*, 2013, Bhattacharya *et al.*, 2009, Carpenter *et al.*, 2011, Field *et al.*, 2009, Gonzalez *et al.*, 2005, Urban *et al.*, 2009).

The β-defensins are a family of multifunctional peptides with roles in inflammation and reproduction as well as direct antiviral and antimicrobial effects (Semple & Dorin, 2012, Dorin & Barratt, 2014, Wiens *et al.*, 2014). In humans, eight β-defensin genes show extensive copy number variation as a block, with a modal copy number of 4 per diploid genome (Hollox *et al.*, 2008). This variation is reflected in levels of β-defensin in the serum, at least for human β-defensin 2 (hbd2), encoded by the *DEFB4* gene (Jansen *et al.*, 2009, Jaradat *et al.*, 2013). Of the eight β-defensin genes, two (*DEFB4* and *DEFB103* encoding hbd2 and hbd3 respectively) have been shown to encode peptides that have anti-HIV activity in vitro (Chang & Klotman, 2004). They have also been shown to have chemotactic activity, and hbd3 has been shown to stimulate the type 1 interferon-β response to the viral ligand mimic polyI:C (Semple *et al.*, 2015).

Some studies have suggested a relationship between low β-defensin copy number and increased HIV susceptibility (Milanese *et al.*, 2009, Mehlotra *et al.*, 2012). However, these studies had small sample sizes and used often unreliable qPCR assays, increasing the potential false positive rate (Mehlotra *et al.*, 2016). Other work, using a more robust triplex paralogue ratio test (PRT) approach to measure β-defensin copy number and much larger sample sizes, examined the association of β-defensin CNV with viral load before initiation of highly-active antiretroviral therapy, and immune reconstitution following initiation of antiretroviral therapy (Hardwick *et al.*, 2012). This study used a cohort of Ethiopian and Tanzanian patients who were naïve to antiretrovirals and were at a late stage of HIV infection (CD4+ T cell count <200 cells/mm^3), and found an association of higher copy number with higher viral load immediately prior to highly-active anti-retroviral therapy (HAART), and with poorer

immune reconstitution. This was in contrast to functional studies that together implied an anti-HIV effect of β-defensins (Quiñones-Mateu *et al.*, 2003, Sun *et al.*, 2006, Sun *et al.*, 2005), but some studies used high concentrations of β-defensins in vitro, often in the absence of serum. At *in vivo* levels, under more realistic assay conditions, the predominant effect of β-defensins at the site of infection may be to act as a chemokine recruiting Th17 cells (Ghannam *et al.*, 2011), which are particularly susceptible to HIV infection (Gosselin *et al.*, 2010, Alvarez *et al.*, 2013).

Given these results, further exploration of the role of the β-defensin CNV in clinical parameters relating to HIV infection was warranted. The IAVI Protocol C (IAVI) is a prospective cohort study that follows HIV progression and transmission longitudinally since seroconversion in a cohort of sub-Saharan Africans from several centres across south and east Africa. The Swiss HIV Cohort study (SHCS) is a longitudinal study of adult HIV patients recruited from across Switzerland where clinical and laboratory parameters are followed at 6-months intervals. Both differ from the initial study in that the viral load at set point (spVL) is the primary clinical variable tested for genetic association, rather than VL immediately prior to HAART or response to HAART (follow up of CD4 count). This is in common with other studies investigating the role of host genetic factors in HIV.

Our aim in this study is to further explore the relationship between β-defensin copy number and HIV infection. We investigate the role of β-defensins in modifying three clinical parameters – HIV susceptibility, HIV VL at setpoint, and HIV progression. We use a method to type β-defensin copy number that has been extensively validated and is considerably more robust than alternative quantitative PCR methods.

Methods

Ethics and cohort details

All participants from the International AIDS Vaccine Initiative (IAVI) and SHCS cohorts were HIV-1 infected adults. The IAVI cohort ~~was~~ comprised ~~of~~ recent HIV-1 seroconverters (SCs) enrolled from Kenya, Rwanda, Uganda, and Zambia between 2006 and 2011, under a uniform study protocol sponsored by IAVI (Amornkul *et al.*, 2013, Price *et al.*, 2011). The procedures for written informed consent and multidisciplinary research activities were approved by institutional review boards at all clinical research centres and participating institutions. SHCS was approved by the local Ethics Committees of all participating centres, and written informed consent was obtained from the participants. The study has enrolled more than 18,000 HIV-infected individuals to date.

Sociodemographic and behavioural data are recorded at entry to the study, in particular year of birth, gender, and the date of the last negative HIV test. Laboratory and clinical data, including viral load and CD4+ T-cell count, are obtained at each semi-annual follow-up visit.

~~B~~ β defensin CNV typing

We used a triplex paralogue ratio test (PRT) for determining diploid copy number at the β-defensin region (Armour *et al.*, 2007, Aldhous *et al.*, 2010, Fode *et al.*, 2011). Briefly, PRT is a form of quantitative PCR where test and reference loci are amplified by the same primer pair minimising the differences in amplification kinetics between them. At the endpoint of PCR, the test and reference products can be distinguished and quantified using capillary electrophoresis. With each PCR, six positive controls of known copy number are used to generate a calibration curve and normalise across experiments. In this study, we used the same six samples throughout, which were the same as used in previous studies, ensuring comparability of data. Data for each cohort were visualised using scatterplots of results from individual assays and histograms of the results from the three assays combined.

Data analysis

spVL was determined previously as the geometric mean of the eligible log10 viral load measurements in each individual (Fellay *et al.*, 2007). Regression models were constructed using the generalised linear model framework in IBM SPSS Statistics v22. Normalised raw PRT copy number (i.e. not rounded ~~n~~or binned) was used as a measure of real underlying copy number of the locus in logistic regression, linear regression and Cox proportional hazards regression models, as ~~used~~ previously (Wain *et al.*, 2014). Covariates in the models were sex, age, and principal components of genetic variation, derived from genomewide SNP genotypes. We repeated all analyses with rounded (binned) integer copy number estimates and maximum-likelihood estimates of integer copy number (Aldhous *et al.*, 2010) with no significant change in results. Integer copy number values presented in Table 1 are from maximum-likelihood analysis calls, consistent with previous publications (Hardwick *et al.*, 2011, Fode *et al.*, 2011, Wain *et al.*, 2014).

Results and Discussion

We typed 387 samples from the IAVI cohort for β-defensin genomic copy number using a previously-published triplex paralogue ratio test (PRT). Clear evidence of clustering of raw copy number was observed, however for the IAVI cohort there was some considerable overlap between copy numbers 3 and 4 (Supplementary Figure 1), which ~~is most likely~~may to be due to heterogeneity between the

multiple locations of sample collection and extraction. Of the 387 samples, 302 had matching spVL data (Table 2). Analysis of the copy number distribution at the 9 different sampling sites showed a modal copy number of 4 or 5, ranging between 1 and 9 (Table 1), broadly consistent with previous data. There is a notable difference in frequency of the 6 copy individuals between the current sample from Lusaka and a previous sample of individuals with unknown HIV status, but this is most likely due to a sampling artefact.

Using the weighted mean raw copy number values generated by PRT, we tested for association with log(spVL) using a generalised linear model, with sex, age, and the first three principal components of genomewide SNP genotype data as covariates. We found no association with β-defensin genomic copy number (β=0.007, 95%CI -0.064 to 0.077, p=0.853, table 3).

We then typed 3155 individuals for β-defensin genomic copy number from the Swiss HIV cohort using triplex PRT. Clustering of raw copy numbers was equivalent to previous studies (Supplementary Figure 1, ~~(~~Hardwick *et al.*, 2012, Hardwick *et al.*, 2011, Wain *et al.*, 2014). Analysis of the copy number distribution showed a modal copy number of 4, ranging between 1 and 9 (Table 1), consistent with previous data. Of these samples, 1525 had matching spVL data (Table 2). Using the weighted mean raw copy number values generated by PRT, we tested for association with ~~spVL usingspVL~~ using a generalised linear model, with sex, age, and the first two principal components of genomewide SNP genotype data as covariates. We found no association with β defensin genomic copy number (β=-0.02, 95%CI -0.06 to 0.021, p=0.335, table 4).

Previous work had also analysed the association of VL with β-defensin copy number by dividing the copy number distribution, ranging from 1 to 9, into two discrete categories, 4 or more copies and fewer than 4 copies. This has the potential to increase power, as a linear response to copy number is not assumed. However, using the same covariates as above, neither the SHCS cohort nor the IAVI cohort showed any association (reference category copy number <4, β=-0.015, 95%CI -0.117 to 0.087, p=0.773 for the SHCS cohort, β=-0.22, 95%CI -0.567 to 0.127, p=0.214 for the IAVI cohort).

In both cohorts we found a highly significant association between males and higher spVL values (Table 3 and Table 4), as has been observed previously (Donnelly *et al.*, 2005, Junghans *et al.*, 1999, Farzadegan *et al.*, 1998). We also found a significant association with the first principal component of genetic variation in the IAVI cohort but not the SHCS cohort (Table 3 and Table 4). However, it should be noted that the first principal component of genetic variation is not comparable between the studies, and in the IAVI dataset it will measure a greater degree of variation across the

geographically-distinct populations sampled, reflecting both genetic and confounding environmental differences between these populations.

An alternative clinical variable which shows evidence of association with host genetic variation is the rate of progression of HIV from seroconversion to a CD4+ T cell level of <350 cells/mm^3 or to treatment start. Progression data for 229 individuals from the SHCS were available. Using a Cox regression model we found no statistically significant association with β-defensin copy number (Exp(B)=1.122, p=0.154, table 5). Similar progression data for the IAVI cohort was available for 301 individuals, and, again, we found no statistically significant association with β-defensin copy number (Exp(B)=0.985, p=0.829, table 6). Both analyses found a statistically significant relationship with initial viral load, as expected, and a lower hazard ratio for women relative to men, reflecting a slower rate of progression in women, as previously observed (Jarrin *et al.*, 2008).

We finally investigated whether there was evidence of association of β-defensin copy number and risk of acquiring HIV by constructing a ~~case-control~~n analysis in which the cases are from the SHCS ~~and the controls are~~are compared with individuals of European descent of unknown HIV status previously typed as part of other studies. ~~For controls, w~~We used 1156 individuals from a population cohort from Nottingham, UK and 695 individuals from Leicester, UK (Wain *et al.*, 2014), combined with 183 UK individuals from the ECACC Human Random Controls cohort (Hardwick *et al.*, 2011). These individuals were of unknown HIV status, and are treated as controls. Using logistic regression with case/control as the binary outcome variable, we found no association with β-defensin copy number (Figure 1, β=0.009, 95%CI -0.042 to 0.061, p=0.725).

Taken together, we find no evidence for association of with HIV susceptibility or spVL. We also find no evidence of a strong effect on HIV progression rate, although it should be noted that the small sample size makes it unlikely that we could detect a small- or medium-sized effect. Recent evidence has shown that common single nucleotide variation at the HLA locus and *CCR5* is responsible for 25% of variability in spVL, and that further studies should be focused on other classes of variation such as rare SNVs and CNVs (McLaren *et al.*, 2015). With increasing affordability of short read sequencing, genomewide analysis of CNV, including complex multiallelic CNVs such as the β-defensin locus, is becoming possible on larger numbers of sequences and ultimately direct genomewide typing of CNV in large cohorts will reveal the contribution to host variation in HIV response, and response to other infectious diseases.

Acknowledgements

Study design: EJH, RA, JF. Data collection and analysis: RA, EJH. Contribution of reagents: PS, PJM, SL, JG, SA, JF. Manuscript preparation: EJH, JF, PS.

We would like to thank the Royal Hashemite Court of Jordan for funding a PhD studentship for RA, Emmanuel Cormier and Matt Price for facilitating the study, and Gurdeep Matharu Lall for technical support.

This study has been partly supported by the Swiss National Science Foundation (Swiss HIV Cohort Study, grant #148522), by SHCS project #651, the SHCS research foundation and NIAID Center for HIV/AIDS Vaccine Immunology (CHAVI) grant AI067854.

Members of the Swiss HIV Cohort Study: Aubert V, Battegay M, Bernasconi E, Böni J, Braun DL, Bucher HC, Burton-Jeangros C, Calmy A, Cavassini M, Dollenmaier G, Egger M, Elzi L, Fehr J, Fellay J, Furrer H (Chairman of the Clinical and Laboratory Committee), Fux CA, Gorgievski M, Günthard H (President of the SHCS), Haerry D (deputy of "Positive Council"), Hasse B, Hirsch HH, Hoffmann M, Hösli I, Kahlert C, Kaiser L, Keiser O, Klimkait T, Kouyos R, Kovari H, Ledergerber B, Martinetti G, Martinez de Tejada B, Marzolini C, Metzner K, Müller N, Nadal D, Nicca D, Pantaleo G, Rauch A (Chairman of the Scientific Board), Regenass S, Rudin C (Chairman of the Mother & Child Substudy), Schöni-Affolter F (Head of Data Centre), Schmid P, Speck R, Stöckle M, Tarr P, Trkola A, Vernazza P, Weber R, Yerly S.

Members of IAVI Protocol C: Price MA, Simek M, Allen S, Karita E, Kilembe W, Lakhi S, Inambao M, Kamali A, Sanders EJ, Anzala O, Edward V, Bekker L-G, Tang J, Gilmour J

IAVI Africa HIV Prevention Partnership:

Kigali, Rwanda (Project San Francisco): Etienne Karita, Principal Investigator; Susan Allen, Principal Investigator; Roger Bayingana, Investigator; Kayitesi Kayitenkore, Investigator

Nairobi, Kenya (Kenya AIDS Vaccine Initiative): Omu Anzala, Principal Investigator; Gaudensia Mutua, Principal Investigator

Kilifi, Kenya (Center for Geographic Medicine Research—Coast & Kenya Medical Research Institute): Eduard J. Sanders, Principal Investigator; Peter Mugo, Investigator

Medical Research Council (MRC)/Uganda Virus Research Institute (UVRI) Uganda Research Unit on AIDS, Entebbe, Uganda: Anatoli Kamali, Principal Investigator; Rogers Twesigye, Study Coordinator, John Byabagambi, Study Physician; Florence Babirye, Nurse

Medical Research Council (MRC)/Uganda Virus Research Institute (UVRI) Uganda Research Unit on AIDS, Masaka, Uganda: Anatoli Kamali, Principal Investigator; Eugene Ruzagira, Study Coordinator, Agnes Bwanika, Ubaldo Bahemuka, and Freddie Mukasa Kibengo, Study Physicians; Peter Hughes, Laboratory Manager; Vincent Basajja, Community Liaison Officer

Lusaka, Ndola and Kitwe, Zambia (Zambia Emory HIV Research Project): William Kilembe, Principal Investigator; Susan Allen, Principal Investigator; Shabir Lakhi, Principal Investigator; Mubiano Inambao, investigator;

Rustenburg, South Africa (Aurum Institute): Mary H. Latka, Principal Investigator; Gavin J Churchyard, Investigator; Petra I Kruger, Investigator; Heeran Makkan, Study Coordinator; Candice M Chetty-Makkan, Study Coordinator; Ben Makhoana, Community Liaison Officer; Tiro Dinake, Nurse; Matsidi Malefo, Senior Research Assistant; Ireen Mosweu, Research Assistant

Cape Town, South Africa (Desmond Tutu HIV Foundation): Linda-Gail Bekker, Principal Investigator; Keren Middelkoop, Investigator; Surita Roux, Investigator

Figure legends

Figure 1 – Cumulative distribution of β-defensin copy number in HIV cases ~~and~~ the general European population ~~controls~~.

Data from the Swiss HIV Cohort study (3155 cases) and UK population (2034 ~~controls~~).

References

Aklillu, E., Odenthal-Hesse, L., Bowdrey, J., Habtewold, A., Ngaimisi, E., Yimer, G., Amogne, W., Mugusi, S., Minzi, O. & Makonnen, E. (2013) CCL3L1 copy number, HIV load, and immune reconstitution in sub-Saharan Africans. *BMC infectious diseases,* 13**,** 536.

Aldhous, M.C., Bakar, S.A., Prescott, N.J., Palla, R., Soo, K., Mansfield, J.C., Mathew, C.G., Satsangi, J. & Armour, J.A. (2010) Measurement methods and accuracy in copy number variation: failure to replicate associations of beta-defensin copy number with Crohn's disease. *Human molecular genetics,* 19**,** 4930-4938.

Alvarez, Y., Tuen, M., Shen, G., Nawaz, F., Arthos, J., Wolff, M.J., Poles, M.A. & Hioe, C.E. (2013) Preferential HIV infection of CCR6+ Th17 cells is associated with higher levels of virus receptor expression and lack of CCR5 ligands. *J Virol,* 87**,** 10843-54.

Amornkul, P.N., Karita, E., Kamali, A., Rida, W.N., Sanders, E.J., Lakhi, S., Price, M.A., Kilembe, W., Cormier, E., Anzala, O., Latka, M.H., Bekker, L.G., Allen, S.A., Gilmour, J., Fast, P.E. & Partnership, I.a.H.P. (2013) Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-Saharan Africa. *Aids,* 27**,** 2775-86.

Armour, J.~~a~~A.L., Palla, R., Zeeuwen, P.L.J.M., Den Heijer, M., Schalkwijk, J. & Hollox, E.J. (2007) Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. *Nucleic acids research,* 35**,** e19-e19.

Bashirova, A.A., Thomas, R. & Carrington, M. (2011) HLA/KIR restraint of HIV: surviving the fittest. *Annual review of immunology,* 29**,** 295-317.

Bhattacharya, T., Stanton, J., Kim, E.-Y., Kunstman, K.J., Phair, J.P., Jacobson, L.P. & Wolinsky, S.M. (2009) Ccl3l1 and hiv/aids susceptibility. *Nature medicine,* 15**,** 1112-1115.

Cantsilieris, S. & White, S.J. (2013) Correlating multiallelic copy number polymorphisms with disease susceptibility. *Human Mutation,* 34**,** 1-13.

Carpenter, D., Walker, S., Prescott, N., Schalkwijk, J. & Armour, J.A. (2011) Accuracy and differential bias in copy number measurement of CCL3L1 in association studies with three auto-immune disorders. *BMC genomics,* 12**,** 418.

Chang, T.L. & Klotman, M.E. (2004) Defensins: natural anti-HIV peptides. *Aids Rev,* 6**,** 161.

Donnelly, C.A., Bartley, L.M., Ghani, A.C., Le Fevre, A.M., Kwong, G.P., Cowling, B.J., Van Sighem, A.I., De Wolf, F., Rode, R.A. & Anderson, R.M. (2005) Gender difference in HIV-1 RNA viral loads. *HIV Med,* 6**,** 170-8.

Dorin, J.R. & Barratt, C.L. (2014) Importance of beta-defensins in sperm function. *Molecular human reproduction,* 20**,** 821-6.

Farzadegan, H., Hoover, D.R., Astemborski, J., Lyles, C.M., Margolick, J.B., Markham, R.B., Quinn, T.C. & Vlahov, D. (1998) Sex differences in HIV-1 viral load and progression to AIDS. *Lancet,* 352**,** 1510-4.

Fellay, J., Shianna, K.V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., Cozzi-Lepri, A., De Luca, A., Easterbrook, P., Francioli, P., Mallal, S., Martinez-Picado, J., Miro, J.M., Obel, N., Smith, J.P., Wyniger, J., Descombes, P., Antonarakis, S.E., Letvin, N.L., Mcmichael, A.J., Haynes, B.F., Telenti, A. & Goldstein, D.B. (2007) A whole-genome association study of major determinants for host control of HIV-1. *Science,* 317**,** 944-7.

Field, S.F., Howson, J.M., Maier, L.M., Walker, S., Walker, N.M., Smyth, D.J., Armour, J.A., Clayton, D.G. & Todd, J.A. (2009) Experimental aspects of copy number variant assays at CCL3L1. *Nat Med,* 15**,** 1115-7.

Fode, P., Jespersgaard, C., Hardwick, R.J., Bogle, H., Theisen, M., Dodoo, D., Lenicek, M., Vitek, L., Vieira, A. & Freitas, J. (2011) Determination of beta-defensin genomic copy number in different populations: a comparison of three methods. *PLOS one,* 6**,** e16768.

Ghannam, S., Dejou, C., Pedretti, N., Giot, J.-P., Dorgham, K., Boukhaddaoui, H., Deleuze, V., Bernard, F.-X., Jorgensen, C. & Yssel, H. (2011) CCL20 and β-defensin-2 induce arrest of human Th17

cells on inflamed endothelium in vitro under flow conditions. *The Journal of Immunology,* 186**,** 1411-1420.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R.J., Freedman, B.I., Quinones, M.P. & Bamshad, M.J. (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science,* 307**,** 1434-1440.

Gosselin, A., Monteiro, P., Chomont, N., Diaz-Griffero, F., Said, E.A., Fonseca, S., Wacleche, V., El-Far, M., Boulassel, M.R., Routy, J.P., Sekaly, R.P. & Ancuta, P. (2010) Peripheral blood CCR4+CCR6+ and CXCR3+CCR6+CD4+ T cells are highly permissive to HIV-1 infection. *Journal of immunology (Baltimore, Md. : 1950),* 184**,** 1604-1616.

Hardwick, R.J., Amogne, W., Mugusi, S., Yimer, G., Ngaimisi, E., Habtewold, A., Minzi, O., Makonnen, E., Janabi, M., Machado, L.R., Viskaduraki, M., Mugusi, F., Aderaye, G., Lindquist, L., Hollox, E.J. & Aklillu, E. (2012) β-defensin Genomic Copy Number Is Associated With HIV Load and Immune Reconstitution in Sub-Saharan Africans. *Journal of Infectious Diseases,* 206**,** 1012-1019.

Hardwick, R.J., Machado, L.R., Zuccherato, L.W., Antolinos, S., Xue, Y., Shawa, N., Gilman, R.H., Cabrera, L., Berg, D.E. & Tyler-Smith, C. (2011) A worldwide analysis of beta-defensin copy number variation suggests recent selection of a high-expressing DEFB103 gene copy in East Asia. *Human Mutation,* 32**,** 743-750.

Hollox, E.J., Barber, J.C.K., Brookes, A.J. & Armour, J.a.L. (2008) Defensins and the dynamic genome: what we can learn from structural variation at human chromosome band 8p23. 1. *Genome Research,* 18**,** 1686-1697.

Hollox, E.J. & Hoh, B.-P. (2014) Human gene copy number variation and infectious disease. *Human Genetics***, 133,** 1217-123317.

Jansen, P.A., Rodijk-Olthuis, D., Hollox, E.J., Kamsteeg, M., Tjabringa, G.S., De Jongh, G.J., Van Vlijmen-Willems, I.M., Bergboer, J.G., Van Rossum, M.M. & De Jong, E.M. (2009) β-Defensin-2 protein is a serum biomarker for disease activity in psoriasis and reaches biologically relevant concentrations in lesional skin. *PLOS one,* 4**,** e4725.

Jaradat, S., Hoder-Przyrembel, C., Cubillos, S., Krieg, N., Lehmann, K., Piehler, S., Sigusch, B. & Norgauer, J. (2013) Beta-defensin-2 genomic copy number variation and chronic periodontitis. *Journal of dental research***,** 00220345135042179292, 1035-1040.

Jarrin, I., Geskus, R., Bhaskaran, K., Prins, M., Perez-Hoyos, S., Muga, R., Hernandez-Aguado, I., Meyer, L., Porter, K. & Del Amo, J. (2008) Gender differences in HIV progression to AIDS and death in industrialized countries: slower disease progression following HIV seroconversion in women. *Am J Epidemiol,* 168**,** 532-540.

Junghans, C., Ledergerber, B., Chan, P., Weber, R., Egger, M. & Study, S.H.CSwiss HIV Cohort Study. (1999) Sex differences in HIV-1 viral load and progression to AIDS. *The Lancet,* 353**,** 589.

Mclaren, P.J., Coulonges, C., Bartha, I., Lenz, T.L., Deutsch, A.J., Bashirova, A., Buchbinder, S., Carrington, M.N., Cossarizza, A., Dalmau, J., De Luca, A., Goedert, J.J., Gurdasani, D., Haas, D.W., Herbeck, J.T., Johnson, E.O., Kirk, G.D., Lambotte, O., Luo, M., Mallal, S., Van Manen, D., Martinez-Picado, J., Meyer, L., Miro, J.M., Mullins, J.I., Obel, N., Poli, G., Sandhu, M.S., Schuitemaker, H., Shea, P.R., Theodorou, I., Walker, B.D., Weintrob, A.C., Winkler, C.A., Wolinsky, S.M., Raychaudhuri, S., Goldstein, D.B., Telenti, A., De Bakker, P.I., Zagury, J.F. & Fellay, J. (2015) Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load. *Proc Natl Acad Sci U S A,* 112**,** 14658-14663.

Mehlotra, R.K., Dazard, J.E., John, B., Zimmerman, P.A., Weinberg, A. & Jurevic, R.J. (2012) Copy Number Variation within Human beta-Defensin Gene Cluster Influences Progression to AIDS in the Multicenter AIDS Cohort Study. *Journal of AIDS & clinical research,* 3, 1000184.

Mehlotra, R.K., Zimmerman, P.A. & Weinberg, A. (2016) Defensin gene variation and HIV/AIDS: a comprehensive perspective needed. *J Leukoc Biol,* 99**,** 687-692.

Milanese, M., Segat, L., Arraes, L.C., Garzino-Demo, A. & Crovella, S. (2009) Copy number variation of defensin genes and HIV infection in Brazilian children. *Journal of acquired immune deficiency syndromes ~~(1999)~~,* 50**,** 331-333.

Pelak, K., Need, A.C., Fellay, J., Shianna, K.V., Feng, S., Urban, T.J., Ge, D., De Luca, A., Martinez-Picado, J., Wolinsky, S.M., Martinson, J., Jamieson, B., Bream, J., Martin, M., Borrow, P., Letvin, N., Mcmichael, A., Haynes, B., Telenti, A., Carrington, M., Goldstein, D., Alter, G. & Immunology", N.C.F.H.a.V. (2011) Copy Number Variation of KIR Genes Influences HIV-1 Control. *PLoS biology,* 9**,** e1001208.

Price, M.A., Wallis, C.L., Lakhi, S., Karita, E., Kamali, A., Anzala, O., Sanders, E.J., Bekker, L.G., Twesigye, R., Hunter, E., Kaleebu, P., Kayitenkore, K., Allen, S., Ruzagira, E., Mwangome, M., Mutua, G., Amornkul, P.N., Stevens, G., Pond, S.L., Schaefer, M., Papathanasopoulos, M.A., Stevens, W., Gilmour, J. & Group, I.E.I.C.S. (2011) Transmitted HIV type 1 drug resistance among individuals with recent HIV infection in East and Southern Africa. *AIDS research and human retroviruses,* 27**,** 5-12.

Quiñones-Mateu, M.E., Lederman, M.M., Feng, Z., Chakraborty, B., Weber, J., Rangel, H.R., Marotta, M.L., Mirza, M., Jiang, B. & Kiser, P. (2003) Human epithelial [beta]-defensins 2 and 3 inhibit HIV-1 replication. *Aids,* 17**,** F39-F48.

Semple, F. & Dorin, J.R. (2012) beta-Defensins: multifunctional modulators of infection, inflammation and more? *J Innate Immun,* 4**,** 337-348.

Semple, F., Macpherson, H., Webb, S., Kilanowski, F., Lettice, L., Mcglasson, S.L., Wheeler, A.P., Chen, V., Millhauser, G.L., Melrose, L., Davidson, D.J. & Dorin, J.R. (2015) Human beta-D-3 Exacerbates MDA5 but Suppresses TLR3 Responses to the Viral Molecular Pattern Mimic Polyinosinic:Polycytidylic Acid. *PLoS Genet,* 11**,** e1005673.

Shea, P.R., Shianna, K.V., Carrington, M. & Goldstein, D.B. (2013) Host genetics of HIV acquisition and viral control. *Annual review of medicine,* 64**,** 203-217.

Sun, L., Demasi, L., Lafferty, M., Goicochea, M., Lu, W. & Garzino-Demo, A. (2006) CCR6 mediates the intracellular HIV inhibitory activity of human beta-defensin 2. *Retrovirology,* 3**,** S77.

Sun, L., Finnegan, C.M., Kish-Catalone, T., Blumenthal, R., Garzino-Demo, P., La Terra Maggiore, G.M., Berrone, S., Kleinman, C., Wu, Z., Abdelwahab, S., Lu, W. & Garzino-Demo, A. (2005) Human beta-defensins suppress human immunodeficiency virus infection: potential role in mucosal protection. *J Virol,* 79**,** 14318-14329.

Urban, T.J., Weintrob, A.C., Fellay, J., Colombo, S., Shianna, K.V., Gumbs, C., Rotger, M., Pelak, K., Dang, K.K., Detels, R., Martinson, J.J., O'brien, S.J., Letvin, N.L., Mcmichael, A.J., Haynes, B.F., Carrington, M., Telenti, A., Michael, N.L. & Goldstein, D.B. (2009) CCL3L1 and HIV/AIDS susceptibility. *Nat Med,* 15**,** 1110-1112.

Wain, L.V., Odenthal-Hesse, L., Abujaber, R., Sayers, I., Beardsmore, C., Gaillard, E.A., Chappell, S., Dogaru, C.M., Mckeever, T., Guetta-Baranes, T., Kalsheker, N., Kuehni, C.E., Hall, I.P., Tobin, M.D. & Hollox, E.J. (2014) Copy number variation of the beta-defensin genes in europeans: no supporting evidence for association with lung function, chronic obstructive pulmonary disease or asthma. *PLOS one,* 9**,** e84192.

Wiens, M.E., Wilson, S.S., Lucero, C.M. & Smith, J.G. (2014) Defensins and viral infection: dispelling common misconceptions. *PLoS Pathog,* 10**,** e1004186.

Formatted: Position: Horizontal: 1.02", Relative to: Page, Vertical: 10.83", Relative to: Paragraph, Width: Exactly 6.32", Height: Exactly 11.69"

391

Formatted: Indent: Left: 0", First line: 0"

Tables

Table 1 – ML integer copy number calls and comparison with other cohorts

Location	β-defensin copy number (frequency)									
	1	2	3	4	5	6	7	8	9	Total
Kigali, Rwanda	0	1 (0.02)	8 (0.12)	31 (0.47)	16 (0.24)	10 (0.15)	0	0	0	66
Masaka, Uganda	0	4 (0.09)	12 (0.26)	15 (0.32)	8 (0.17)	2 (0.04)	3 (0.06)	1 (0.02)	1 (0.02)	47
Kilifi, Kenya	0	5 (0.2)	4 (0.16)	12 (0.48)	3 (0.12)	1 (0.04)	0	0	0	25
Kangemi, Kenya	0	0	1 (0.11)	2 (0.22)	5 (0.55)	1 (0.11)	0	0	0	9
Lusaka, Zambia	1 (0.01)	9 (0.10)	15 (0.16)	30 (0.32)	30 (0.32)	3 (0.03)	4 (0.04)	1 (0.01)	1 (0.01)	94
Entebbe, Uganda	0	0	1 (0.08)	7 (0.54)	2 (0.15)	2 (0.15)	1 (0.08)	0	0	13
Copperbelt, Zambia	0	3	10 (0.24)	13 (0.32)	11 (0.27)	3 (0.07)	1 (0.02)	0	0	41
Rustenberg, South Africa	0	0	1 (0.17)	2 (0.33)	3 (0.5)	0	0	0	0	6
Cape Town, South Africa	0	0	0	1	0	0	0	0	0	1
Lusaka, Zambia (Hardwick	0	3 (0.03)	16 (0.13)	44 (0.37)	24 (0.20)	25 (0.21)	6 (0.05)	1 (0.01)	1 (0.01)	120

et al., 2011)										
SHCS	15 (0.00)	118 (0.04)	575 (0.18)	1326 (0.42)	808 (0.26)	243 (0.08)	54 (0.02)	13 (0.00)	3 (0.00)	3155
Combined European controls	11 (0.01)	76 (0.04)	390 (0.19)	833 (0.41)	484 (0.24)	191 (0.09)	39 (0.02)	6 (0.00)	4 (0.00)	2034

Table 2 **Descriptive statistics of the two cohorts analysed in this study**

	Mean (sd) SHCS cohort	**Mean (sd) IAVI cohort**
n	1525	302
Sex	1236 (81%) male, 289 (19%) female	178 (59%) male, 124 (41%) female
Year born or age	1964.84 (10.503)	31.7 (8.48)
β-defensin copy number	4.26 (1.05)	4.18 (1.14)
Log(spVL) copies/mL	4.36 (0.87)	4.4 (0.76)

Table 3 **Regression model for association with Log(spVL) for IAVI cohort**

Variable	**B**	**95% Wald Confidence Interval for B**		**p-value**
		Lower	Upper	
Sex: (reference=female)	0.371	0.196	0.546	$3.4x10^{-5}$
Age (years)	0.004	-0.006	0.015	0.392
Genetic variation principal component 1	3.076	1.599	4.554	$4.5x10^{-5}$
Genetic variation principal component 2	-1.584	-3.077	-0.090	0.038
Genetic variation principal component 3	1.094	-0.459	2.647	0.167
β-defensin copy number	0.007	-0.064	0.077	0.853

B=regression coefficient

Table 4 Regression model for association with Log(spVL) for SHCS cohort

Variable	B	95% Wald Confidence Interval for B		p-value
		Lower	Upper	
Sex: (reference=female)	0.439	0.330	0.548	2.8×10^{-15}
Age (years)	0.002	-0.002	0.006	0.270
Genetic variation principal component 1	0.498	-1.650	2.646	0.650
Genetic variation principal component 2	-0.357	-2.473	1.760	0.741
β-defensin copy number	-0.020	-0.060	0.021	0.335

Table 5 Cox regression analysis of time-to-death outcome in SHCS cohort

Variable	B	Standard error of B	Wald-statistic	p-value	Exp(B)
Year of Birth	0.004	0.010	0.187	0.668	1.004
Sex (1=male, 2=female)	-0.574	0.249	5.327	0.021	0.563
Genetic variation principal component 1	2.210	4.923	0.202	0.653	9.115
Genetic variation principal component 2	-3.521	5.453	0.417	0.519	0.030
Log (spVL)	0.669	0.117	32.717	1.1×10^{-8}	1.953

β-defensin copy number	0.115	0.081	2.036	0.154	1.122

The Wald statistic is a measure of the departure of B from zero, calculated as the square of the estimate of B divided by the square of the estimated standard error of B, will have a chi-squared distribution, and is used to calculate the p-value. Exp(B) represents the Hazard ratio.

Table 6 Cox regression analysis of time-to-death outcome in IAVI cohort

Variable	B	SE	Wald-statistic	p-value	Exp(B)
Year of Birth	-0.019	0.010	3.510	0.061	0.982
Sex (1=male, 2=female)	-0.487	0.182	7.110	0.008	0.615
Genetic variation principal component 1	-0.903	1.531	0.348	0.555	0.405
Genetic variation principal component 2	-3.659	1.581	5.358	0.021	0.026
Genetic variation principal component 3	1.661	1.878	0.783	0.376	5.267
Log (spVL)	1.222	0.151	65.055	$7.3x10^{-16}$	3.393
β-defensin copy number	-0.015	0.068	0.047	0.829	0.985

The Wald statistic is a measure of the departure of B from zero, calculated as the square of the estimate of B divided by the square of the estimated standard error of B, will have a chi-squared distribution, and is used to calculate the p-value. Exp(B) represents the Hazard ratio.

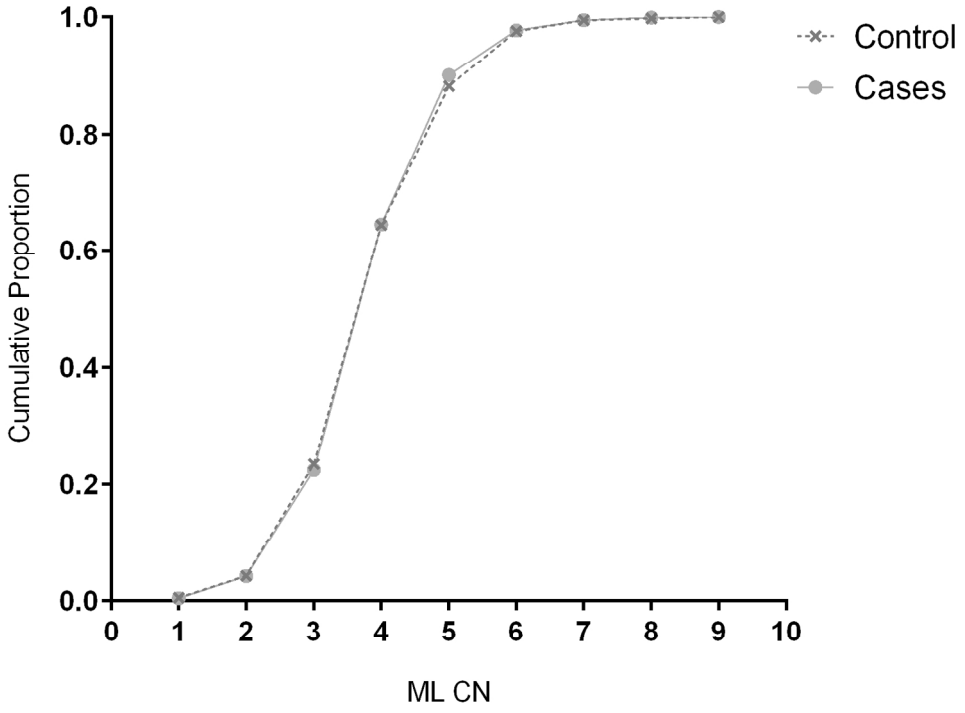
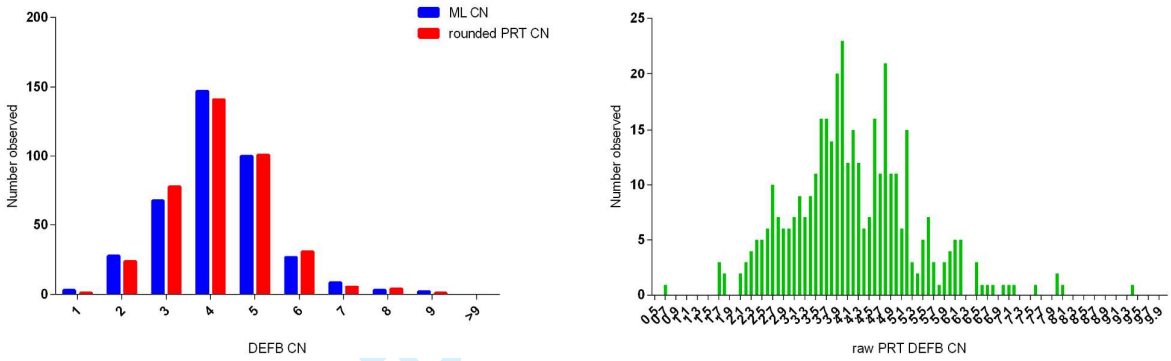


Figure 1 – Cumulative distribution of β-defensin copy number in HIV cases and population controls. Data from the Swiss HIV Cohort study (3155 cases) and UK population (2034 controls).

Supplementary figure 1

Beta-defensin copy number distributions of (a) IAVI cohort and (b) SHCS. The bar graphs on the left shows the distribution of integer copy numbers for each cohort calculated by two different approaches (maximum-likelihood and rounding). The histograms on the right show the distribution of the raw normalised copy numbers.

a)



b)