BRIEF REPORT

Evidence After Imputation for a Role of *MICA* Variants in Nonprogression and Elite Control of HIV Type 1 Infection

Sigrid Le Clerc,[1] Olivier Delaneau,[4] Cédric Coulonges,[1] Jean-Louis Spadoni,[1] Taoufik Labib,[1] Vincent Laville,[1] Damien Ulveling,[1] Josselin Noirel,[1] Matthieu Montes,[1] François Schächter,[1] Sophie Caillat-Zucman,[2,3] and Jean-François Zagury[1]

[1]Chaire de Bioinformatique, EA4627, Conservatoire National des Arts et Métiers, [2]Hôpital Robert Debré AP-HP, and [3]UMR 1149 INSERM-Université, Paris, France; and [4]Département de Génétique et Développement, Faculté de Médecine, Université de Genève, Geneva, Switzerland

Past genome-wide association studies (GWAS) involving individuals with AIDS have mainly identified associations in the *HLA* region. Using the latest software, we imputed 7 million single-nucleotide polymorphisms (SNPs)/indels of the 1000 Genomes Project from the GWAS-determined genotypes of individuals in the Genomics of Resistance to Immunodeficiency Virus AIDS nonprogression cohort and compared them with those of control cohorts. The strongest signals were in *MICA*, the gene encoding major histocompatibility class I polypeptide-related sequence A ($P = 3.31 \times 10^{-12}$), with a particular exonic deletion ($P = 1.59 \times 10^{-8}$) in full linkage disequilibrium with the reference *HCP5* rs2395029 SNP. Haplotype analysis also revealed an additive effect between *HLA-C*, *HLA-B*, and *MICA* variants. These data suggest a role for *MICA* in progression and elite control of human immunodeficiency virus type 1 infection.

Keywords. AIDS; elite control; GWAS; HIV-1; indel; imputation; MICA; non progression; SNP.

More than 15 genome-wide association studies (GWAS) have been published regarding AIDS progression [1]. They have mainly pointed to associations in the *HLA* region of chromosome 6, involving candidate functional variants in the HLA-B groove [2, 3] and the *HLA-C* region [4]. Our group has also performed a GWAS using the Illumina 317K beadchip on the Genomics of Resistance to Immunodeficiency Virus (GRIV) cohort, the largest cohort of nonprogressor patients in the world, and we also identified the *HCP5* rs2395029 single-nucleotide polymorphism (SNP) corresponding to *HLA-B57*01* [3]. All of these past GWAS relied on DNA chips that covered only a subset of genomic SNPs (up to 2.5 million for the latest Illumina chips) but did not address important markers, such as indels. In the present study, we computationally reanalyzed the genotypic data found in the GRIV cohort by means of Shape-IT and IMPUTE 2 software [5, 6], exploiting the 40 million markers (SNPs and indels) from the recently released 1000 Genomes Phase I integrated variant set [7].

METHODS

Samples

The GRIV Cohort

The GRIV cohort was established in 1995 with white individuals of European descent living in France to identify host genes associated with nonprogression to AIDS. Nonprogressors were characterized as individuals who had had asymptomatic HIV type 1 infection for >8 years, had not received treatment, and had a $CD4^+$ T-cell count of >500/mm^3. DNA was obtained from fresh peripheral blood mononuclear cells or from Epstein-Barr virus–transformed cell lines. All patients provided written informed consent before enrollment in the GRIV study.

The nonprogressor group (n = 266) was composed of 197 males and 69 females aged 19–62 years at the time of study inclusion. The viral load (plasma HIV RNA concentration) at inclusion could be assessed for 248 individuals and allowed us to define a group of elite controllers: 49 had a viral load of <100 copies/mL. For haplotype analysis, we distinguished the elite controllers from the remaining nonprogressors, hereafter referred to as "nonelite nonprogressors" (n = 217).

The DESIR Control Group

This control group has been previously described by Le Clerc et al [8] and comprised 694 nonobese, normoglycemic, HIV-1–seronegative French participants from the DESIR trial. A total of 281 individuals were male, and 413 were female, and ages ranged from 30 to 64 years. All patients signed an informed consent before enrollment in the genetic association study.

Additional Control Groups

In addition to the DESIR genotype data, we also used the genotypes from individuals in the control cohort Supplémentation en

Received 12 February 2014; accepted 3 June 2014; electronically published 16 June 2014.
Correspondence: Jean-François Zagury, MD, PhD, 292 rue Saint Martin, 75003 Paris, France (zagury@cnam.fr).

The Journal of Infectious Diseases® 2014;210:1946–50

DOI: 10.1093/infdis/jiu342

Vitamines et Minéraux Antioxydants (SU.VI.MAX; n = 1352) to confirm the allelic frequencies of the main SNPs identified in our case-control study. This control cohort of seronegative subjects was previously described by Limou et al [3].

We also used the rapid progressors from the GRIV cohort (n = 84) and the seroconverters from the ACS cohort (n = 404) as supplementary control groups, to ensure that the association results were not linked to infection. These 2 latter cohorts have been described and reviewed by Limou et al [1].

Genotyping and Quality Control

Specimens from the GRIV nonprogressors and the control groups were genotyped with the Illumina Infinium II HumanHap300 BeadChips (Illumina, San Diego, CA). The methods and quality controls were described by Le Clerc et al [8] and Limou et al [3].

SNP Imputation

For each population, we started from the original genotypes obtained by chips [1]. Our pipeline worked with a prephasing step, using Shape-IT [5], followed by imputation using IMPUTE2 [6]. This pipeline has been shown to be optimal [6]. To improve accuracy and keep a reasonable number of subjects in each group, some groups were prephased together, as follows: GRIV nonprogressors and rapid progressors together with the DESIR controls (n = 1044), the SU.VI.MAX controls (n = 1352), the ACS controls (n = 404). SNP genotypes were imputed using the 1000 Genomes Phase I integrated variant set [7] (with SNPs and indels) as the reference panel. Only the variants imputed with high reliability (imputation quality score $P > .8$ [9]) were retained (n = 7 172 846).

Haplotypes

Haplotypes were computed with the Shape-IT software [5] for the following 3 SNPs: the reference SNP rs9264942, corresponding to *HLA-C*; the reference SNP rs2395029, corresponding to the *HCP5*/*HLA-B*5701*/*MICA*017* variant; and the SNP rs112243036, corresponding to the strongest association found in *MICA* (the gene encoding major histocompatibility class I polypeptide-related sequence A) by the present study. The haplotypes were computed after grouping together the controls, the elite controllers, and the nonelite nonprogressors, and they were then extracted separately for each subpopulation.

Statistical Analysis

The approximately 7 million reliably imputed polymorphisms were compared between the nonprogressor group (n = 266) and the DESIR control group (n = 694). For each SNP, phenotype was regressed on genotype dose by logistic regression with the SNPtest software [9] in the additive mode, taking into account stratification by adding the 2 first Eigenstrat PC axes as covariates. For haplotype analysis, we compared the group of elite controllers (n = 49) and nonelite nonprogressors (n = 217) with the control group (n = 694) using the Fisher exact test.

RESULTS AND DISCUSSION

After imputation based on the 317 000 SNPs initially genotyped in the 266 GRIV nonprogressor cohort and the 694 DESIR controls, we obtained reliable genotypes for 7 million variants. We compared the nonprogressors from the GRIV cohort to the DESIR controls (all of whom were white individuals) with respect to these 7 million variants and found 55 signals passing genome-wide significance ($P < 5 \times 10^{-8}$) in the chromosome 6 *HLA* region only, in agreement with the previous GWAS on HIV outcomes. Of those, 48 are in high linkage disequilibrium ($r^2 > .8$) with the previously described *HCP5* rs2395029 SNP (the tagSNP for the allele *HLA-B*5701*). A majority of those SNPs/indels are located within the *HLA-B* and *MICA* genes (Figure 1*A*), with the most significant *P* values (up to 3.31×10^{-12}) observed for the *MICA* gene, compared with 3×10^{-9} for the *HLA-B* SNPs (Supplementary Table 1). Our strongest signal ($P = 3.31 \times 10^{-12}$) corresponded to the *MICA* intronic rs112243036 SNP already identified for viral control by the International HIV Controllers Study [2]. Whereas the function of the *HLA-B* association has been extensively investigated, the *MICA* association has not. We also looked at the best SNP frequencies in several additional cohorts (rapid progressors from the GRIV cohort, seroconverters from the ACS cohort, and seronegative controls from the SU.VI.MAX cohort) and found them to be similar to those in the DESIR cohort (Supplementary Table 2), confirming that the associations observed were indeed linked with nonprogression to AIDS and not with susceptibility to HIV infection.

As shown in Supplementary Table 1, among the 55 signals passing the Bonferroni threshold, 48 are located close to or within a gene (ie, near the 3′ or 5′ regions, in introns, in exons, or in the 3′ untranslated region), the location and the genetic context of the identified SNPs are shown in Figure 1*A*. Among the 4 exonic variants, the most remarkable was the deletion rs199503730 ($P = 1.59 \times 10^{-8}$) in the *MICA* gene, whose functional impact is likely to be more dramatic than the impact of the 2 nonsynonymous mutations (Supplementary Table 1). Indeed, the deletion induces a reading frameshift located just before the transmembrane section of the MICA protein, leading to a variation of the protein sequence and a longer hydrophobic region in the transmembrane region (Figure 1*B*). This deletion is carried only by the 2 alleles *MICA*017* and *MICA*015* [10]. Since *MICA*015* is extremely rare in the European population and has always been observed in associated with *HLA-B*45* [11], rather than with the *HLA-B*5701* allele found in our patients, we will only discuss *MICA*017* in the remainder of this article. Interestingly, in our cohort, the *MICA* deletion (and, thus, *MICA*017*) was in perfect linkage disequilibrium with the reference *HCP5*

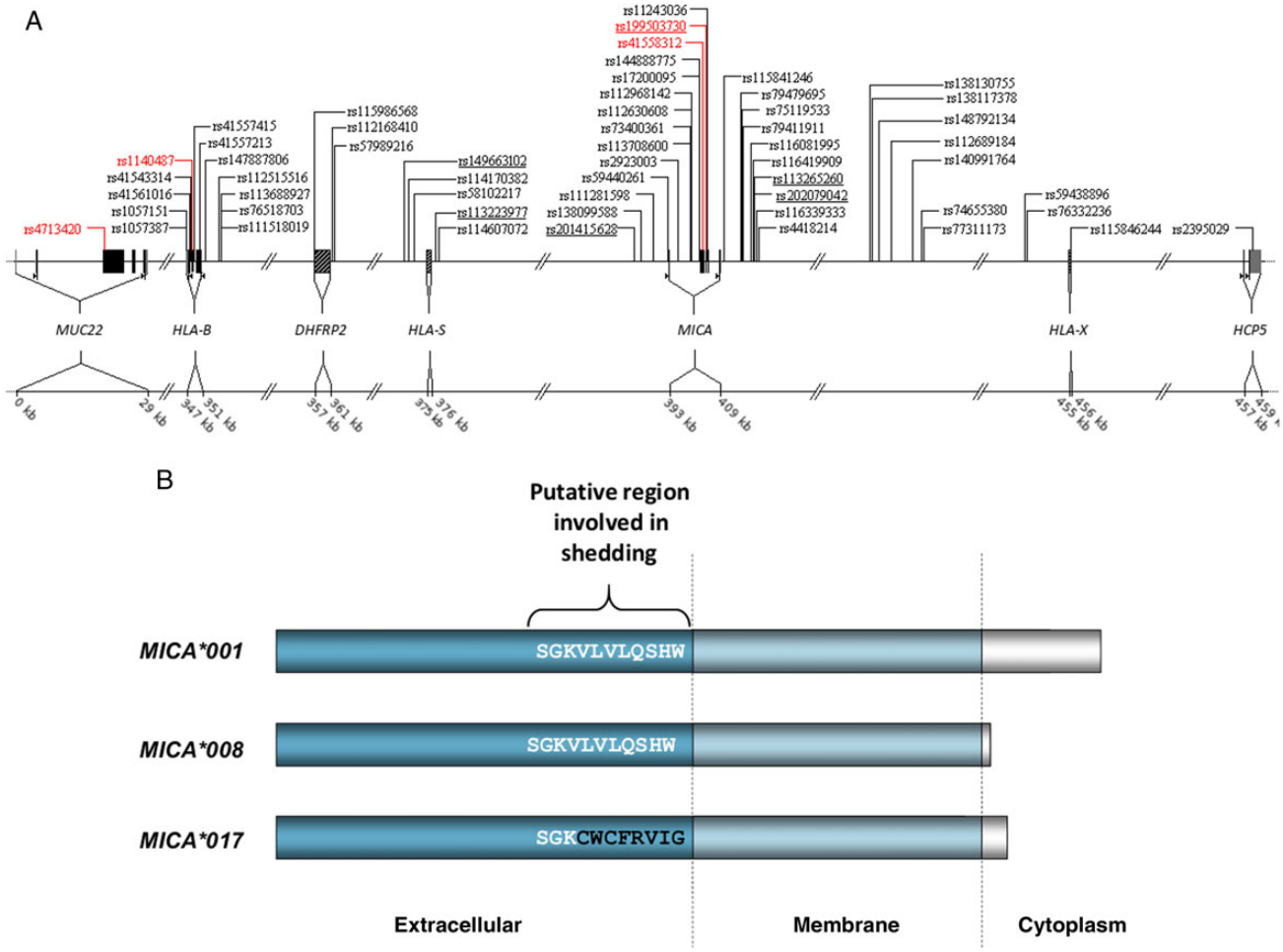

Figure 1. *A*, Genetic map of the *HLA* region. Only the single-nucleotide polymorphisms found significant after imputation in this study are represented. The exonic variants are in red, the other variants are in black, and the indels are underlined. The black boxes are the exons, the gray boxes represent the 3′ and 5′ untranslated regions, and the striped boxes in black and white represent pseudogenes. The start and stop codons for each gene are black arrows in the direction of the transcription. The scale is given in kilobases (kb). *B*, Representation of the major histocompatibility class I polypeptide-related sequence A (MICA) proteins coded by the alleles *MICA*001* (the reference *MICA* allele), *MICA*008* (the most frequent allele in Europeans), and *MICA*017* (corresponding to the deletion rs199503730). The sequence variations before the transmembrane region, which corresponds to the putative cleavage site involved in the shedding of MICA, are detailed. The *MICA*017* sequence is unique among the *MICA* alleles of the European population.

rs2395029 SNP and with *HLA-B*5701* ($r^2 = 1$). Since it is in the vicinity of the putative cleavage zone of soluble MICA [12], this deletion may impact the shedding of MICA (Figure 1*B*), thus providing an additional functional explanation for the well-known genetic association of the rs2395029 SNP.

In addition to *HLA-B*5701*, previous works had also unraveled associations between *HLA-C* variant rs9264942 and potential biological explanations of viral control and nonprogression [1, 3]. To see whether our *MICA* associations could shed new light on these associations, we evaluated the haplotypes of the most significant variants of *HLA-C, HLA-B* and *MICA* in controls (n = 694), nonelite nonprogressors (n = 217), and elite controllers (n = 49): rs9264942, for *HLA-C* [1]; rs2395029, for *HLA-B*5701* [1] (in full linkage disequilibrium with the rs199503730 deletion corresponding to *MICA*017*); and rs112243036, for *MICA* (the strongest signal identified by our study and previously mentioned by Pereyra et al [2]; Supplementary Table 1). Analysis of the haplotypes allowed us to draw some conclusions about the interplay between rs9264942, rs2395029, and rs112243036 (Table 1). In contrast to what has been described in the literature [2], the frequency of haplotype CTG (odds ratio [OR], 1.00; 95% confidence interval [CI], 0.62–1.60; $P = 1$) suggested that the *HLA-C* rs9264942 C allele was not effective unless combined with the *HLA-B*5701*/*MICA*017* rs2395029 G allele and/or the *MICA* rs112243036 A allele for viral control and nonprogression (haplotypes CGG and CTA, respectively; Table 1). A major signal was also observed for haplotype CGA (OR, ∞; 95% CI, 6.03–∞; $P = 2.65 \times 10^{-4}$)

Table 1. Analysis of Haplotype Associations for the 3 Reference Single-Nucleotide Polymorphisms (SNPs)

Haplotype			Allelic Frequency (%)			Elite vs Control		NPNE vs Controls	
HLA-C	*HLA-B*5701*	*MICA*	Control (n = 694)	Elite (n = 49)	NENP (n = 217)	*P* Value	OR (95% CI)	*P* Value	OR (95% CI)
T	T	G	54.90	25.00	39.90	1.05×10^{-8}	0.27 (.16–.44)	6.87×10^{-8}	0.54 (.43–.68)
C	T	G	31.30	31.20	30.30	1	1.00 (.62–1.60)	7.65×10^{-1}	0.96 (.75–1.22)
C	T	**A**	6.20	18.70	15.00	5.90×10^{-5}	3.47 (1.87–6.19)	5.98×10^{-8}	2.66 (1.85–3.80)
T	T	**A**	4.80	9.40	6.80	5.46×10^{-2}	2.06 (.87–4.34)	1.06×10^{-1}	1.45 (.89–2.33)
C	**G**	**A**	0.00	3.10	0.20	2.65×10^{-4}	∞ (6.03–∞)	2.35×10^{-1}	∞ (.08–∞)
T	**G**	G	0.70	4.20	2.30	1.00×10^{-2}	5.95 (1.34–21.15)	1.31×10^{-2}	3.29 (1.22–8.88)
C	**G**	G	2.30	8.30	5.20	3.20×10^{-3}	3.83 (1.48–8.82)	4.94×10^{-3}	2.29 (1.26–4.13)
T	**G**	**A**	0.00	0.00	0.20	NaN	NaN	2.36×10^{-1}	∞ (.08–∞)

rs9264942 was analyzed for *HLA-C*, rs2395029 was analyzed for *HLA-B*5701* (in full linkage disequilibrium with rs199503730 deletion corresponding to *MICA*017*), and rs112243036 corresponded to the best SNP located in *MICA*. The linkage disequilibrium between the 3 SNPs is weak, with r^2 vales of <0.1 between each SNP for each population. The minor alleles of interest for each variant are highlighted in bold.

Abbreviations: CI, confidence interval; Elite, elite controllers; NaN, not a number; NENP, nonelite nonprogressors; OR, odd ratio.

when comparing elite controllers to controls, whereas the signal became weaker when comparing nonelite nonprogressors to controls, suggesting that these 3 SNPs were more important for viral control than for the nonprogression phenotype. The frequency of this latter haplotype was confirmed independently in other cohorts of elite controllers (C. Coulonges, personal communication). We used logistic regression to determine whether the combined effect of the 3 SNPs was additive or synergistic by comparing the 2 models by means of analysis of variance. The effect appeared to be additive, as the improvement brought about by the inclusion of the interaction between all 3 SNPs was not significant ($P = .10$).

Overall, this novel analysis of imputed SNPs/indels suggests a prominent role for variants of the MICA protein with multiple associations found in this gene.

First, the strongest associations were observed for *MICA* variants. *MICA* encodes a protein that functions as a ligand for NKG2D, which is present on $CD8^+$ α/β T cells, γ/δ T cells, and natural killer (NK) cells [13]. The engagement of NKG2D activates cytolytic responses of γ/δ T cells and NK cells against infected cells and tumor cells [13].

Second, the *MICA* deletion rs199503730, which corresponds to the alleles *MICA*017* and *HLA-B*5701* in the white population, was consistently overrepresented among elite controllers in our study (15.32%), as well as those in other published studies (13.9% in the study by Pereyra et al [2] and 16.7% in the study by Dalmasso et al [14]), compared with controls. This *MICA* deletion is located near the putative cleavage site of MICA, ahead of the transmembrane region [12], and could thus lead to a lower level of soluble MICA by preventing shedding (Figure 1*B*). Recent studies have measured higher levels of soluble MICA for individuals with chronic HIV infection, compared with healthy controls and HIV controllers [15], which could be important since soluble MICA interacts with the NKG2D receptor and impairs the cytotoxicity of NK cells during HIV infection [15]. This shedding would not occur in the *HLA-B*5701/MICA*017* elite controllers: the exceptional strength of the well-known *HCP5* rs2395029 association with nonprogression and viral load control could thus be explained by the additive functional effects of the *HLA-B*5701* and *MICA*017* variants.

Third, the haplotype analysis of the elite controller and nonelite nonprogressor groups revealed a combined effect of the well-known *HLAB*5701/MICA*017* allele, *HLA-C* rs9264942-C allele, and the strongest *MICA* signal, rs112243036 A allele. This latter effect could be explained by the combined action of MICA on innate immunity and of HLA on adaptive immunity.

The use of imputation has unraveled new signals in the *MICA* gene with a putative biological explanation, and our haplotype analysis revealed the combined effects of *HLA-C*, *HLA-B*, and *MICA* variants in the GRIV cohort. All of the data suggest a role of *MICA* in disease progression and elite control and underline the importance of the interaction of several components of the immune system during AIDS progression. These results provide new clues for the mechanisms of elite control and nonprogression of HIV infection and will require confirmation by functional investigations.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://jid.oxfordjournals.org). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. We thank all the patients and medical staff who have kindly collaborated with the GRIV project.

Financial support. This work was supported by the Agence Nationale de Recherche sur le SIDA et les Hépatites Virales, the AIDS Cancer Vaccine Development Foundation, Peptinov, and the Conservatoire National des Arts et Métiers (fellowship to S. L. C.).

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Limou S, Zagury JF. Immunogenetics: genome-wide association of non-progressive HIV and viral load control: HLA genes and beyond. Front Immunol **2013**; 4:118.
2. International HIVCS, Pereyra F, Jia X, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science **2010**; 330:1551–7.
3. Limou S, Le Clerc S, Coulonges C, et al. Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). J Infect Dis **2009**; 199:419–26.
4. Apps R, Qi Y, Carlson JM, et al. Influence of HLA-C expression level on HIV control. Science **2013**; 340:87–91.
5. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods **2012**; 9:179–81.
6. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet **2012**; 44:955–9.
7. Genomes Project C, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. Nature **2010**; 467:1061–73.
8. Le Clerc S, Limou S, Coulonges C, et al. Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for AIDS (ANRS Genomewide Association Study 03). J Infect Dis **2009**; 200:1194–201.
9. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet **2007**; 39:906–13.
10. Obuchi N, Takahashi M, Nouchi T, et al. Identification of MICA alleles with a long Leu-repeat in the transmembrane region and no cytoplasmic tail due to a frameshift-deletion in exon 4. Tissue Antigens **2001**; 57:520–35.
11. Gao X, Single RM, Karacki P, Marti D, O'Brien SJ, Carrington M. Diversity of MICA and linkage disequilibrium with HLA-B in two North American populations. Hum Immunol **2006**; 67:152–8.
12. Kaiser BK, Yim D, Chow IT, et al. Disulphide-isomerase-enabled shedding of tumour-associated NKG2D ligands. Nature **2007**; 447:482–6.
13. Bauer S, Groh V, Wu J, et al. Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. Science **1999**; 285:727–9.
14. Dalmasso C, Carpentier W, Meyer L, et al. Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. PLoS One **2008**; 3:e3907.
15. Nolting A, Dugast AS, Rihn S, et al. MHC class I chain-related protein A shedding in chronic HIV-1 infection is associated with profound NK cell dysfunction. Virology **2010**; 406:12–20.