HHS Public Access

Author manuscript

Infect Genet Evol. Author manuscript; available in PMC 2018 April 01.

Published in final edited form as:

Infect Genet Evol. 2017 April ; 49: 78–87. doi:10.1016/j.meegid.2017.01.007.

RICH2 is implicated in viraemic control of HIV-1 in black South African individuals

Maria Paximadis[1,2,*], **Refilwe N. Ngqobe**[1], **Richard E. Chaisson**[3], **Neil A. Martinson**[4], and **Caroline T. Tiemessen**[1,2]

[1]National Institute for Communicable Diseases, Centre for HIV and STIs, Cell Biology, Johannesburg, South Africa

[2]School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[3]Johns Hopkins University Center for AIDS Research, Baltimore, MD, USA

[4]Perinatal HIV Research Unit (PHRU), University of the Witwatersrand, Johannesburg, South Africa and MRC Soweto Matlosana Centre for HIV/AIDS and TB Research, South Africa

Abstract

An intronic single nucleotide polymorphism (SNP) in *RICH2* (rs2072255; 255^{i}), in complete linkage disequilibrium (LD) with an exonic SNP (rs2072254; 254^{e}), has been identified in a genome wide association study to be associated with progression to AIDS in Caucasian individuals. RICH2 links tetherin to the cortical actin network and the RICH2/tetherin interaction has been shown to be important for the downstream activation of NF-κβ and the consequential promotion of proinflammatory responses. We investigated the role of these two SNPs in natural control of HIV-1 in black South Africans including healthy controls (HCs; N=102) and antiretroviral-naive HIV-1-infected controllers (HICs; N=52) and progressors (N=74). HICs were stratified as elite controllers (ECs; N=11), viraemic controllers (VCs; N=30), high viral load (VL) long term non-progressors (HVL LTNPs; N=11) and also according to VL<400 RNA copies/ml (HICs VL<400; N=20) and VL>400 RNA copies/ml (HICs VL>400; N=32). Results showed that in contrast to Caucasians who had very strong LD between these SNPs (r^2=0.97), black populations exhibited low LD (r^2=0.11–0.27), however a 254^{e} minor allele was always present with a 255^{i} minor allele but not vice versa. The SNPs did not show significant over- or underrepresentation in any particular group, however the combination of 254^{e} major allele homozygosity and 255^{i} heterozygosity (254^{e}AA/255^{i}GA) was underrepresented in HICs (OR=3.26; *P*=0.04) and VCs (OR=7.77; *P*=0.02) compared to HCs, and in HICs VL>400 compared to both HCs (*P*=0.002) and progressors (*P*=0.02). A lower $CD4^{+}$ T-cell count was associated with 254^{e}AA/255^{i}GA and 255^{i} (GA+AA) in the total HIV-1-infected group (*P*=0.043) and progressors (*P*=0.017), respectively. *In silico* analysis predicted loss of an exonic splice

[*]Correspondence: Maria Paximadis, Centre for HIV and STIs, Cell Biology, National Institute for Communicable Diseases, 1 Modderfontein Road, Sandringham, 2131, Johannesburg, South Africa., paxim@nicd.ac.za, Tel: +27 11 386 6365.

enhancer site with the 254[e]-G allele. We postulate that altered splicing of RICH2 will affect levels of RICH2 expression and consequently NF-κβ activation. These findings point to a role for RICH2 and tetherin in viraemic natural control of HIV-1.

Keywords

HIV-1; natural control; host genes; RICH2; tetherin

1. Introduction

HIV-1-infected individuals have variable rates of disease progression and viral control. On one extreme of the spectrum we find individuals with failure to control and rapid disease progression, usually within three years of seroconversion, and on the opposite extreme are a group of rare individuals (approximately 1 in 300) termed elite controllers (ECs), who are able to control virus to less than 50 RNA copies/ml, without antiretroviral (ARV) treatment (Walker, 2007). However, between these two extremes we find individuals with a broad range of disease progression and viral control and the description and categorisation of these individuals tends to vary greatly between different studies (Casado et al., 2010; Gurdasani et al., 2014). 'Healthy' $CD4^+$ T-cell counts (generally >500 cells/μl whole blood), a low viral load set point (VLS; <2000 RNA copies/ml) and prolonged ARV-free infection (generally ≥7 years) are the criteria most often used to define HIV-1 controllers (HICs). There are however HICs that do not fit all criteria, for example HIV-1-infected individuals that have a low VLS (usually <2000 RNA copies/ml) with sustained high $CD4^+$ T-cell counts. Although not all these individuals have documented prolonged follow up, or know their date of HIV infection, a low VLS predicts slow disease progression, and these individuals remain asymptomatic for many years (Mellors et al., 1995). We refer to these individuals as viraemic controllers (VCs) in this study. There are also HIV-1-infected individuals who maintain high $CD4^+$ T-cell counts (generally >500 cells/μl) for prolonged periods without ARV treatment despite high viral loads (generally >10 000 RNA copies/ml), a similar phenotype to SIV-infected sooty mangabeys who preserve their $CD4^+$ T-cells despite high viral loads and experience no AIDS-related events (Hunt, 2009; Okulicz et al., 2009; Rotger et al., 2011; Silvestri et al., 2007). In this study we refer to these individuals as high viral load long term non-progressors (HVL LTNPs).

The study of individuals displaying these broad and variable clinical phenotypes provides valuable insights into the biology and mechanisms of viral control and disease progression. The large variability in HIV-1 control has been attributed to: host genes variability, viral pathogenicity and environmental factors (Fellay et al., 2009). In a recent genome wide association study (GWAS), up to 25% of the variability in viral load was attributed to common genetic variations in the Human Leukocyte Antigen (HLA) and CCR5-encoding regions of the genome (McLaren et al., 2015). Many individuals that control infection however do not possess these well described host gene variants. This suggests undiscovered protective mechanisms and pathways. Genome wide association studies detect common variants and rarer gene variants, potentially responsible for a large fraction of unexplained inter-individual variation may be missed (Fellay et al., 2009). Furthermore, most GWAS

studies have focussed on HIV-1-infected populations of European ancestry (Caucasian) infected with subtype B HIV-1 while the burden of HIV-1 disease lies with subtype C infected individuals in sub-Saharan Africa (WHO, 2015). Populations of different ethnic origins differ greatly with respect to genetic profiles and linkage disequilibrium patterns (Campbell and Tishkoff, 2008) and immunogenetic findings from one population group cannot simply be extrapolated to another ethnically distinct population group. Thus, functionally important minor genetic variants may have been missed with the failure to extensively study populations most affected by the HIV-1 epidemic. Validating Caucasian-based HIV-1 related gene-association findings in sub-Saharan African population groups is therefore imperative.

A GWAS study set up to specifically screen for low frequency [minor allele frequency (MAF) <5%] single nucleotide polymorphisms (SNPs) associated with HIV-1 disease identified a new risk allele (intronic SNP rs2072254-A allele) for progression to AIDS in the *RICH2* gene (Le Clerc et al., 2011). This study, conducted on Caucasian cohorts, included subtype B HIV-1-infected slow, and rapidly progressing HIV-infected individuals. RICH2 (also known as ARHGAP44), is a Rho-type GTPase activator protein (Rho-GAP) and was originally described as a homologue of RICH1 (also known as Nadrin and ARHGAP17) (Furuta et al., 2002). Proteins containing a Rho-GAP domain usually function to catalyze the hydrolysis of GTP that is bound to Rho, Rac and/or Cdc42 (regulators of the actin cytoskeleton), thereby inactivating these regulators (Peck et al., 2002). Interestingly, in polarised cells, RICH2 has been shown to link tetherin (also known as BST2) to the cortical actin network (Rollason et al., 2009). In a more recent study (Galao et al., 2014), RNAi depletion of RICH2 *in vitro* blocked detection of phosphorylated tetherin which resulted in failure of tetherin to induce NF-κβ, however it did not inhibit tetherin's physical antiviral function (i.e. inhibition of release of enveloped virus particles from infected cells). This association of RICH2 with a well-described anti-HIV-1 factor (tetherin) offers a possible explanation for the role of the *RICH2*-associated SNP in HIV-1 disease progression.

To our knowledge, there have been no reports describing the prevalence of this *RICH2* SNP in black South Africans. Le Clerc et al. (2011) report the *RICH2* rs2072255 SNP to be in complete linkage disequilibrium (LD) with another *RICH2* SNP, namely rs2072254, which they suggest was located in a splicing site of *RICH2*, thereby making this SNP the likely causal SNP explaining the association. In this study we assessed the role of these two *RICH2* SNPs in HIV-1 control by determining their representation in black ARV-free HIV-1-infected controllers displaying a broad range of control phenotypes, progressors and in HIV-1 negative black control individuals.

2. Materials and Methods

2.1 Study Populations

Three black South African (SA) groups were included in this study. The first study group, termed healthy controls (HCs), are HIV-1-uninfected individuals (N=102), a subset of the previously described ESKOM (Electricity Supply Commission) cohort which included individuals working at sites of this South African company countrywide and represent a sample of the black South African population (Paximadis et al., 2012).

The second study group is comprised of 52 HIV-1-infected controllers, termed HIV controllers (HICs) and were recruited from sites in Johannesburg and Soweto. These HICs are a heterogeneous group of HIV-1 infected adults able to control HIV-1 disease and were included on the basis of one or a combination of the following: persistently high $CD4^+$ T-cell count (>500 cells/μl whole blood), documented duration of infection without ARV therapy (ART) and viral load (VL, HIV-1 RNA copies/ml plasma). The HICs were categorised into three subgroups: ECs (N=11) were selected on basis of at least one VL test yielding <50 RNA copies/ml and were infected for an unspecified length of time without ART and had $CD4^+$ T-cell counts >500 cells/μl; VCs (N=30) had VL<2000 RNA copies/ml in the absence of ARVs with $CD4^+$ T-cell counts >500 cells/μl, and were also infected for an unspecified length of time; HVL LTNPs (n=11) were selected on basis of VL>10 000 RNA copies/ml for a period of ≥7 years and were not on ARVs and had no $CD4^+$ T-cell count below 500 cells/μl. For purpose of analysis we also stratified and analysed the HICs on their VL alone, having VL<400 RNA copies/ml (HICs VL<400; N=20) and VL>400 RNA copies/ml (HICs VL>400; N=32). A VL of 400 RNA copies/ml was selected in our study to differentiate between high and low viral replication due to one of our VCs having a reported VL of <400 RNA copies/ml at the time point of analysis.

The third study group is comprised of 74 HIV-1-infected black SA progressors recruited from Chris Hani Baragwanath Hospital, Soweto. These individuals were recruited based on a decline of $CD4^+$ T-cell from above 500 cells/μl to <350 cells/μl and were initiated on ART with a VL>10 000 RNA copies/ml at time of ART initiation. This group is predominantly represented by individuals progressing at a typical or normal rate (median time since HIV-1 diagnosis to ART initiation is 6 years).

Additional information on the HIV-1-infected individuals, including the median length of time since HIV-1 diagnosis, is shown in Table 1. HIV-1-infected individuals recruited from Chris Hani Baragwanath Hospital, Soweto formed part of a larger cohort previously described (Hoffmann et al., 2016; Martinson et al., 2014).

HIV-1 RNA levels were quantified using the COBAS® AmpliPrep/COBAS® Taqman® HIV-1 test, v1.5 and v2.0 ultrasensitive tests (Roche Diagnostic Systems,Inc., New Jersey, USA) and the FACSCount System (Becton Dickinson, San Jose, California, USA) was used to determine $CD4^+$ T-lymphocyte counts.

This study was approved by the University of the Witwatersrand Committee for Human Research Ethics Committee (Medical) and written informed consent was obtained from all individuals participating in this study.

2.2 Genotyping of the RICH2 rs2072254 and rs2072255 SNPs

PCR primers were designed using the *RICH2* sequence (NCBI: dbSNP using either rs2072255 or rs2072254) to amplify and sequence a short region (450bp) of the gene harbouring the two SNPs of interest from a small subset of our HCs as well as from Caucasian volunteers (in order to increase the odds of finding all three possible genotypes). Since the two SNPs are found in close proximity (68 base pairs apart; Figure 1A), we amplified a single fragment to cover the region harbouring the two SNPs. Genomic DNA

was extracted from whole blood using QIAamp DNA Blood Mini Kit (Qiagen), according to the manufacturer's guidelines. The PCR primers used were as follows: *RICH2* forward 5′-GTTACTGAGCCAGTGTGATGAGC-3′ and *RICH2* reverse 5′-CTTGTAGCCACACAGCCTCC-3′. PCR reactions were carried out using the EXPAND High Fidelity PCR System (Roche Applied Science, Switzerland) using 300 nM final concentration of each primer, 200 μM final concentration of each dNTP and 1.5 mM final concentration $MgCl_2$. Cycling conditions were the standard conditions for EXPAND High Fidelity PCR System (see manufacturer's instructions). Purified amplicons were then sequenced using the same primers used for the PCR. Sequencing reactions were set up Applied Biosystems Big Dye Terminator v3.1 chemistry (Life Technologies, Thermo Fisher Scientific, Massachusettes, USA) and electrophoresed on an Applied Biosystems 3100 Genetic Analyser (Life Technologies). The resulting chromatograms were then analyzed using the Sequencher software (Gene Codes Corp., Ann Arbor, MI USA).

Using gDNA from control individuals harbouring the various genotypes, we then developed SYBR Green real-time C_T (cycle threshold) shift PCR assays to detect the two SNPs using allele-specific PCR with two allele-specific primers (designed with their 3′-end bases complementary to each of the two SNP alleles present) and one common reverse or forward primer (depending on orientation of allele-specific primers). This method has been previously described (Paximadis et al., 2013). Primer sequences for C_T shift assays were as follows, for the rs2072254 SNP assay, reverse C-allele-specific primer, 5′-AGGGGGCTACTCGGAAGAG[C]-3′, reverse T-allele-specific primer, 5′-AGGGGGCTACTCGGAAGAG[T]-3′ and common forward primer, 5′-GATGTTGTACCTTGGAGTTG-3′; for the rs2072255 SNP assay, reverse T-allele-specific primer, 5′-AACATCAAAGCGCAGAGACC[T]-3′, reverse C-allele-specific primer, 5′-AACATCAAAGCGCAGAGACC[C]-3′ and common forward primer, 5′-GTTGGGCAAATGCATGTAAGAG-3′. Square brackets denote lock-nucleic-acid (LNA) modified nucleotides. Real-time reactions were set up in 96-well plates. Each 10 μl reaction contained 1×Fermentas SYBR Green PCR Master Mix (Thermo Fisher Scientific, Massachusettes, USA), 10 pmol of each forward and reverse primer, and ~10–50 ng gDNA template. The PCR reactions were run in an Applied Biosystems 7500 Real-Time PCR system. An initial denaturation (10 minutes) at 95°C was followed by 40 cycles of denaturation (15 seconds) at 95°C, annealing (30 seconds) at 60°C and extension (1 minute) at 72°C.

2.3 Linkage disequilibrium and test for Hardy-Weinberg equilibrium

The Haploview version 4.2 (Barrett et al., 2005) software was used to calculate linkage disequilibrium (LD) and deviation from Hardy-Weinberg (HW) equilibrium for the two *RICH2* SNPs genotypic data generated for our study populations. In addition, the genotypic data was also visually studied to detect linkage patterns that were not obvious using LD calculation alone. We also accessed individual genotypic data for the two SNPs available on the 1000 genomes project browser (The Genomes Project, 2015) for the total African population (N=661), the total European population (N=503) and the sub-Saharan Yoruba population from Nigeria (N=108). LD between the two *RICH2* SNPs was also calculated for

these populations using Haploview version 4.2 and compared to our black SA population (HCs).

2.4 In silico analysis of the two RICH2 SNPs

Le Clerc et al. (2009) suggested the exonic rs2072254 SNP to be the functional SNP due to FASTSNP (Yuan et al., 2006) analysis predicting it to fall within a *RICH2* splice site. We further carried out a thorough investigation of both SNPs using a range of online *in silico* splice site-related analysis tools (Table S2). Given that the two SNPs are only 68 nucleotides apart and that they lie in close proximity and on either side of the *RICH2* intron 10/exon 11 junction (Figure 1A), we analysed these two SNPs in the following three contexts:

- **i.** Predicted effect of SNPs (alone and in combination) on the acceptor splice site that should be present at the intron 10/exon 11 junction.
- **ii.** Predicted effect of intronic rs2072255 SNP with respect to the introduction of alternate splice sites.
- **iii.** Predicted effect of exonic rs2072255 SNP on exonic splicing regulation.

2.5 Data Analysis

Frequencies of the alleles of the two SNPs genotyped for all the cohort individuals described above were determined by direct counting. To compare representation of the individual SNPs (allelic and genotypic) as well as combinations of the two SNPs between the relevant study groups, we used Fisher's exact tests implemented through the online statistical package *SISA* (Simple Interactive Statistical Analysis; (Uitenbrook, 1997)). Two-sided tests were used and the level of statistical significance for all analyses was set at $P<0.05$. Odds ratios (OR) and 95% confidence intervals (CI) of the ORs were also calculated using *SISA*. Given the exploratory nature of this work we did not apply correction for multiple comparisons across all comparisons; however we did test our most significant finding by applying Bonferroni correction for number of hypotheses tested.

The SNPs were also analysed in the context of markers of disease progression, namely VL, $CD4^+$ T-cell count and rate of $CD4^+$ T-cell decline using the Mann-Whitney U test implemented through GraphPad Prism version 5.03 for Windows (GraphPad Software, San Diego California USA, www.graphpad.com). The HICs and progressors were analysed as a total group as well as two separate groups for VL and $CD4^+$ T-cell count and only progressors were analysed for rate of $CD4^+$ T-cell decline, where $CD4^+$ T-cell counts were determined every 6 months for a maximum period of 24 months. For this analysis not all progressors had longitudinal $CD4^+$ T-cell counts (N=20) and were thus excluded from this part of the analysis. To calculate rate of $CD4^+$ T-cell count decline, the earliest $CD4^+$ T-cell count from this monitoring period was subtracted from the last $CD4^+$ T-cell count (prior to initiation of ART) and divided by the relevant months to calculate a $CD4^+$ T-cell decline per month. Two additional progressors were excluded from this analysis. One was an outlier with a decline rate value of 97 $CD4^+$ T-cells/month, where the median $CD4^+$ T-cell decline rate was 5.6 $CD4^+$ T-cells/month and the mean $CD4^+$ T-cell decline rate was 8.3 $CD4^+$ T-cells/month (including outlier). The second excluded progressor had two $CD4^+$ T-cell counts of 6 and 3 $CD4^+$ T-cells/μl six months apart.

3. Results

3.1 Representation of the rs2072254 and rs2072255 RICH2 SNPs in the black South African population

Allele data generated for the two *RICH2* SNP positions (Supplementary Figure S1) were tested for deviation from HW equilibrium. Each group of individuals was tested separately. There was no significant deviation from HW equilibrium for both SNP positions in all study groups tested.

The allelic frequencies (AF) and genotypic frequencies (GF) of the two *RICH2* SNPs were calculated for the HCs and compared to data available for these two SNPs, for select populations, from the 1000 Genomes project. The AF and GF for black HCs are shown in supplementary Table S1. The MAF for both *RICH2* SNPs in black HCs (rs2072254: 1.96%; rs2077255: 14.7%) were comparable to those for the sub-Saharan Yoruba population (YRI) from Nigeria (rs2072254: 4.6%; rs2077255: 14.8%) and the total African population (rs2072254: 3.3%; rs2077255: 13.6%). The European MAFs differed from the three African MAFs with a higher rs2072254 SNP MAF (7.7%) and a lower rs2077255 SNP MAF (7.5%).

The representation of the various genotypes for the two SNPs in the HCs and three reference 1000 Genomes populations are shown graphically in Figure 1. For the intronic rs2077255 SNP, heterozygosity (GA) in the HCs was present at relatively high frequency (25.5%), and had very similar representation to both the YRI (25.9%) and total African (24.2) groups, with the European population having close to half the representation (13.3%). For the exonic rs2077254 SNP however, heterozygous (AG) representation in the HCs was only 3.9% compared to 9.3% in the YRI and 6.2% in the total African population groups. The European population had 13.7% heterozygous rs2077254 SNP representation, very similar to the heterozygous rs2077255 SNP representation, suggesting strong LD as reported by Le Clerc et al. (2011).

3.2 Linkage Disequilibrium between rs2072254 and rs2072255 RICH2 SNPs

The calculated LDs between the two *RICH2* SNPs in the black SA HCs as well as the three reference 1000 Genomes populations are shown in Figure 1. The European population had almost complete LD with an r^2=0.97 (D′=1), with only 2/503 individuals being heterozygous for the rs2072254 SNP (AG) and homozygous for the rs2072255 SNP major allele (GG). The total African population and the YRI population showed low LD with r^2 values of 0.20 (D′=0.97) and 0.27 (D′=1), respectively and the black SA HCs had the lowest LD between SNPs (r^2=0.11; D′=1). It was interesting that the D′ values in the African populations (black SA HCs, YRI and total African) suggested strong LD given the low r^2 values and the evident low LD when visually studying the data. There was however an interesting linkage pattern between these two SNPs in these population groups, in that minor allele at the exonic rs2072254 SNP position is always found with a minor allele at the intronic rs2072255 SNP position, however the relationship is not exclusive in the opposite direction, i.e. a minor allele at the rs2072255 SNP position can, and often does occur without a corresponding mutation at the rs2072254 SNP position. This suggests that LD

between these two SNPs is conditional or one-directional and more interestingly may represent a functional relationship between these two SNPs.

Furthermore, when we determined the LD between these two SNPs in the HIV-1-infected study groups, we found that LD varied considerably between these groups, with progressors having low LD (r^2=0.25; D′=1) similar to YRI and total African populations and the HICs as a group having moderate LD (r^2=0.51; D′=1). When the HICs were stratified according to VL though, the HICs VL>400 showed high LD (r^2=0.78; D′=1) compared to the HICs VL<400 with low LD (r^2=0.29; D′=1). These results suggested a disproportionate distribution of these two SNPs as a pair in the various study groups which was verified upon further analysis.

3.3 The individual effect of the rs2072255 and rs2072254 RICH2 SNPs on HIV-1 control

The AF and GF of the two *RICH2* SNPs in all the study groups are listed in supplementary Table S1. Supplementary Figures S2A and S2B show a graphic representation of the MAFs for the rs2072254 and the rs2072255 SNPs, respectively, in all study groups. Comparison of minor AF and GF between HCs and HICs and HIC subgroups and well as progressors and HICs and HIC subgroups are shown in Table 2. Although Figure S2A shows HCs, and to some degree progressors, having lower representation of the exonic rs2072254 SNP minor allele compared to HICs and in particular HICs with higher VLs, comparison of the AF and GF for this SNP did not reveal significant associations (Table 2). There were however trends of higher representation of the rs2072254 minor allele (G) and heterozygous genotype (AG) in the total HICs compared to HCs (OR=0.3; *P*=0.09 for both) and in VCs compared to HCs (G: OR=0.28; *P*=0.08, AG: OR=0.27; *P*=0.08). In addition the HICs VL>400 had trends of higher representation of heterozygosity (AG) for the rs2072254 and lower representation of heterozygosity (GA) for the rs2072255 SNP compared to HCs (OR=0.29; *P*=0.09 and OR=3.28; *P*=0.08, respectively).

3.4 Combined effect of the two RICH2 SNPs rs2072254 and rs2072255 on HIV-1 control

Since the two *RICH2* SNPs showed an interesting one-directional linkage pattern we assessed the combined effect of the various genotypes of these two SNPs with respect to HIV-1 control. Table 3 shows the representation of the various genotypic combinations found in the study groups. Only five of the possible nine combinations were detected. The absence of minor allele homozygosity for both SNPs (rs2072254 GG/rs2072255 AA) was not surprising since this is likely to be a rare occurrence. The three combinations that involve a minor allele at the exonic rs2077254 SNP position (254^e) and no corresponding minor allele at the intronic rs2077255 SNP position (255^i) were not detected in any of the study groups (i.e. 254^eAG/255^iGG; 254^eGG/255^iGA and 254^eGG/255^iGG). This was in agreement with the observation that a mutation at the rs2077254 SNP position was always found with a corresponding mutation at the rs2077255 SNP position.

We compared representation of the two most prevalent genotypic combinations (254^eAA/255^iGA and 254^eAG/255^iGA), other than homozygosity for the major allele at both positions (254^eAA/255^iGG) between study groups. Figure 2A shows results of comparisons between study groups and Figure 2B shows the representation of all genotype

combinations in the HCs, HICs VL<400, HICs VL>400 and progressor groups. The combination of major allele homozygosity at the exonic SNP rs2077254 position and heterozygosity at the intronic SNP rs2077255 position (254^{e}AA/255^{i}GA) was significantly underrepresented in HICs compared to HCs (OR=3.26; P=0.04) and this was also significantly underrepresented in VCs compared to HCs (OR=7.77; P=0.02). Furthermore when we stratified HICs according to VL less and greater than 400 RNA copies/ml, we found that 254^{e}AA/255^{i}GA was completely absent in the HICs VL>400 group (N=32) compared to a 22.6% representation in HCs (P=0.002; with Bonferroni correction P=0.07) and a 16.2% representation in progressors (P=0.02). Since the HICs VL>400 group included the HVL LTNP group with VL>10 000 RNA copies/ml, we compared the HICs VL>400 excluding the HVL LTNPs (i.e. N=21) to both HCs and progressors (data not shown). Significance was maintained compared to HCs (P=0.01) and a strong trend was still evident compared to progressors (P=0.06), however although associations were weaker prior to exclusion of the HVL LTNPs, they were stronger compared to the VC group comparisons. Thus, this *RICH2* SNP genotype combination appears to be deleterious with respect to HIV-1 control in individuals with higher viraemia. Heterozygosity at both these SNP positions (254^{e}AG/255^{i}GA) was marginally elevated in HICs (9.62%) compared to HCs (2.94%) and VCs showed a trend of higher representation of 254^{e}AG/255^{i}GA compared to HCs (OR=0.25; P=0.09).

3.5 RICH2 SNPs and markers of disease progression

When we analysed the two *RICH2* SNPs individually as well as the combined genotypes in the context of markers of disease progression, no effect was seen with respect to VL. Within the progressor group however there was a significant (P=0.017) lower $CD4^{+}$ T-cell count associated with heterozygosity and/or homozygosity for the minor allele of the 255^{i} SNP compared to homozygosity for the major allele (Figure 3A). When we analysed the same SNP for its effect on the rate of $CD4^{+}$ T-cell decline in the progressor study group (Figure 3B), although not significant (P=0.17), there was a higher rate of $CD4^{+}$ T-cell decline associated with the same 255^{i} SNP alleles (i.e. heterozygosity and/or homozygosity for the minor allele compared to homozygosity for the major allele). The 254^{e}AA/255^{i}GA genotype combination of the two *RICH2* SNPs that was significantly underrepresented in HICs with VL>400 RNA copies/ml was also found to be significantly (P=0.04) associated with lower $CD4^{+}$ T-cell count compared to individuals homozygous for the major alleles when the total HIV-1-infected group was analysed (Figure 3C), i.e. HICs and progressors together, even though the $CD4^{+}$ T-cell counts distinctly fell into two groups (>500 and <500 $CD4^{+}$ T-cells/μl). In the progressor group alone, when individuals with the 254^{e}AA/255^{i}GA genotypes were compared to remaining progressors (i.e. with 254^{e}AA/255^{i}GG or 254^{e}AG/255^{i}GA), there was a strong trend (P=0.08) of lower CD4+ T-cell count associated with the 254^{e}AA/255^{i}GA progressors (Figure 3D). No significant associations or trends were noted when analysing the combined *RICH2* SNP genotypes in relation to rate of $CD4^{+}$ T-cell decline.

3.6 Functional implications of the two RICH2 SNPs

The 255^{i} SNP is located 65 nucleotides upstream from the start of exon 11 in the *RICH2* gene and the 254^{e} SNPis located at the third nucleotide position of the codon encoding for

the first amino acid of exon 11 resulting in a synonymous Gly→Gly substitution. Results of all *in silico* analyses conducted to predict the functional implications of these two SNPs are listed in Table 4. When we analysed the effect of the two SNPs on the predicted 3′ acceptor splice site for intron 10 (score for255^{i}WT/254^{e}WT in Table 4), four out of the six analysis tools predicted a score change for at least one of the SNPs, with three programmes predicting a lowered score due to the 254^{e} SNP and one predicting an increased score due to this same SNP. Only one analysis tool (SplicePort) predicted a lower score due to the 255^{i} minor allele and the combination of minor alleles at both SNP positions (255^{i}mt/254^{e}mt) resulted in the lowest predicted score by SplicePort for the acceptor splice site (2.04 compared to 2.12, Table 4).

We also investigated the individual effects of the two SNPs, with respect to either introducing alternate splice sites in the case of the 255^{i} SNP, or altering or introducing an exonic splicing regulator site in the case of the 254^{e} SNP. Three analysis tools predicted a change by the minor allele of the 255^{i} SNP with one tool (ASSP) predicting the introduction of an alternative cryptic donor site, with a fairly good score, and SplicePort and HSF both predicting the introduction of new acceptor splice sites. Interestingly these two acceptor splice sites overlap in sequence and are thus likely to represent the same splice site, with SplicePort having a fairly low predictive score (hence less likely) and HSF having a relatively strong predictive score of 71.7 (range 0–100).

The most consistent prediction however was related to the analysis of the 254^{e} SNP with respect to exonic splice regulation. All four analysis tools predicted the loss of an exonic splice enhancer (ESE) site in the presence of the minor allele (Table 3). Serine/arginine-rich splicing factor 2 (SRSF2), also called splicing factor SC35, was predicted by ESEfinder, AASsites and HSF to be the ESE site lost, whereas the SKIPPY tool did not specifically name the ESE site that is lost.

4. Discussion

Individuals infected with HIV-1 vary greatly in their ability to control their infection in the absence of treatment, irrespective of ethnic background. Those able to control HIV-1 also fall within a broad spectrum of control and it is probable that the mechanisms responsible for controlling HIV-replication and disease are complex, multi-factorial and involve different arms of both the innate and adaptive immune system in the various “categories” of control. The majority of studies that have been carried out to better understand correlates of protective immunity and the role of host genetics on the rate of HIV-1-associated disease progression have predominantly included Caucasian populations and have been essential to the understanding of how host genetics influence control of HIV-1 and disease progression; however populations of different ethnic backgrounds vary substantially in their genetic variation (SNPs, indels, gene copy numbers and haplotype structures), making it necessary to test these findings in populations of sub-Saharan African origin that account for almost 70% of global new HIV-1 infections (WHO, 2015). Furthermore, given that African populations harbour the greatest human genetic diversity, gene-based studies conducted in these populations offers opportunities for discovery of new disease-related factors (Ramsay et al., 2011).

Tetherin (also called BST-2) is a type II transmembrane protein with a unique topology that allows it to tether enveloped viruses to the surface of infected cells, thereby restricting virus release, and in addition it is able to induce innate immune responses through the induction of cytokine/chemokine expression via the activation of NF-κβ (Galao et al., 2012); (reviewed (Mahauad-Fernandez and Okeoma, 2016)). Although HIV-1 can counteract the effect of tetherin through its Viral protein U (Vpu), the neutralizing activity by Vpu on tetherin does not appear to be absolute (Homann et al., 2011) suggesting that the balance of tetherin and Vpu may determine an individual's ability to restrict viral dissemination, and thus may be a key molecule involved in the natural control of HIV-1. RICH2 (ARHGAP44), a BAR-RacGAP protein, has been shown to link tetherin to the actin cytoskeleton by providing a physical link between the cytosolic domain of tetherin and EBP50 (ERM-binding phosphoprotein 50) and ezrin (Rollason et al., 2009). The RICH2/tetherin interaction has been shown to be important for the downstream activation of NF-κβ and the consequential promotion of proinflammatory responses, however depletion of RICH2 by RNAi does not interfere with tetherin's physical antiviral (tethering) function (Galao et al., 2014). One can thus postulate that changes in RICH2 expression levels and/or physical structure could affect NF-κβ activation and the extent of the proinflammatory response.

It was thus of interest that a GWAS study-identified SNP in *RICH2*'s intron 10 (rs2072255; 254^{i}) was found to be associated with subtype B HIV-1 disease progression in Caucasian individuals (Le Clerc et al., 2011). The minor allele of this *RICH2* 254^{i} SNP was found to be significantly underrepresented in slow progressors compared to healthy HIV-1-uninfected individuals in the dominant mode (i.e. heterozygosity + homozygosity for minor allele). Rapid progressing individuals and healthy individuals had similar high allele representation (17.1% and 18.87, respectively) compared to slow progressing individuals (9.04%). Le Clerc et al. (2011) postulated that a downstream exonic SNP (rs2072254; 254^{e}) in LD with the 254^{i} SNP was likely to be the causative SNP due to *in silico* prediction (FASTSNP;(Yuan et al., 2006)) placing it in a splicing site. We thus investigated the role of these two *RICH2* SNPs in natural control of HIV-1 in black individuals, who bear the major burden of HIV-1 infection in South Africa.

Allelic and genotypic frequencies of these two *RICH2* SNPs in black healthy uninfected controls were comparable to those of the Yoruba (Nigeria) and total African populations available on the 1000 Genomes project (The Genomes Project, 2015) but differed significantly from the 1000 Genomes project European (Caucasian) population (MAFs P=0.001 for both SNPs). Furthermore the LD pattern between these two SNPs differed between blacks and Caucasians. Although not in complete LD as reported by Le Clerc et al. (2011), the Caucasian population group had very strong LD (r^2=0.97) between these two SNPs, whereas all black population groups analysed (1000 Genomes total African and Yoruba; SA HCs, HICs and progressors), had low LD (r^2=0.11–0.27) with an interesting one-directional linkage pattern where a minor allele at the 254^{e} SNP site was consistently accompanied by a minor allele at the 255^{i} SNP site (but not vice versa). This relationship was indicative of a potential functional relationship between these two closely located SNPs.

Given the very strong LD between these two SNPs in Caucasian populations, association studies alone cannot differentiate these two SNPs in terms of function; however the low LD

in our black SA population offers the opportunity to interrogate the role of each SNP on HIV-1 control independently. Both allelic and genotypic frequencies of both SNPs in HIV-1-infected controllers, progressors as well as healthy controls revealed no significant differences between groups and corresponding subgroups, however, it was interesting to note that the 254^{e} SNP had strong trends of higher representation of the minor allele (G) as well as heterozygosity (AG) in the HICs and VCs compared to HCs. In addition, when stratifying controllers according to VL, the HICs with higher VLs (>400 RNA copies/ml) showed strong trends of higher representation of heterozygosity for the 254^{e} SNP (OR=0.29) and lower representation of heterozygosity for the 255^{i} SNP (OR=3.28) compared to HCs, thus appearing to have opposing effects. Similarly to the Le Clerc et al. (2011) study, the shifts were seen with respect to HCs, and rapid progressors and HCs had similar allelic and genotypic frequencies. What is different in this study is that the progressor group, with a 6 year median since HIV-1 diagnosis likely represents individuals that have progressed at a typical or normal rate as opposed to rapid or slow.

When representation of the combined genotypes at the two *RICH2* SNP positions was compared between controllers, progressors and healthy controls, it became evident that rather than either of the two SNPs individually affecting HIV-1 control, it was the combination of homozygosity for the 254^{e} SNP major allele together with 255^{i} SNP heterozygosity (254^{e}AA/255^{i}GA) that was significantly lower in HICs and VCs compared to HCs. Moreover, when HICs were stratified according to VL, HICs VL>400 RNA copies/ml had zero representation of 254^{e}AA/255^{i}GA compared to 22.5% representation in HCs and 16.2% representation in progressors. When conservative Bonferroni correction for multiple comparisons was applied, the HCs vs. HICs VL>400 association (P=0.002) maintained a strong trend (P=0.07). These results suggest that in individuals that are able to naturally control HIV-1 in the presence of higher viral replication, the RICH2 protein, likely through its association with tetherin, is playing a significant role, and that the 254^{e}AA/255^{i}GA genotype combination is deleterious in this scenario. In another GWAS study that identified a SNP in the *CXCR6* gene to be associated with natural control of HIV-1, it was only after exclusion of the ECs from the larger controller cohort that the association became apparent and was thus attributed with viraemic long term non-progression and not elite control (Limou et al., 2010). Limou et al. (2010) used a VL of 100 HIV-1 RNA copies/ml of plasma as their stratification between ECs and the remaining controllers. When we excluded our ECs from our total controller group the 254^{e}AA/255^{i}GA combination remained significantly underrepresented in this group compared to HCs (2.4% vs. 22.5% respectively; OR=10.56, P=0.0045), furthermore when we used 100 HIV-1 RNA copies/ml as our stratification value, the 254^{e}AA/255^{i}GA combination maintained significant underrepresention in controllers with VL>100 HIV-1 RNA copies/ml of plasma compared to HCs (2.6% vs. 22.5% respectively; OR=9.94, P=0.008). Thus both the CXCR6 study (Limou et al., 2010) and current study point to different mechanisms being at play in elite control and in the ability to control in the presence of detectable replicating virus. When we analysed these *RICH2* SNPs with respect to CD4+ T-cell count and VL, known markers of disease progression, 254^{e}AA/255^{i}GA was significantly associated with lower CD4+ T-cell count compared to homozygosity for the major alleles (254^{e}AA/255^{i}GG) in the total HIV-1-infected group (HICs + progressors). The 255^{i} SNP alone (dominant mode) was also associated with

significant lower CD4+ T-cell count in the progressor group. Interestingly, no effect was seen on VL with either of the two SNPs or combinations thereof. The preservation of $CD4^+$ T-cells in the presence of replicating virus is reminiscent of SIV-infected sooty mangabeys who manage to have high $CD4^+$ T-cell count and do not succumb to disease despite very high levels of viral replication (Silvestri et al., 2007). From the complete LD between the two SNPs in the Le Clerc el al. (2011) Caucasian study population, we can assume that the 254^eAG/255^iGA + 254^eGG/255^iAA combinations (equivalent to 255^i GA+AA) would have been deleterious, predominantly 254^eAG/255^iGA since minor allele homozygosity for both SNPs would have been present in low frequency (0.8% representation in the 1000 Genomes European population), whereas the 254^eAA/255^iGA combination would not have been detected in that population group. In this study, the 254^eAG/255^iGA combination however showed a trend of higher representation in VCs compared to HCs (OR=0.25), and hence advantageous and showing opposing results to the Le Clerc el al. (2011) study. Furthermore Le Clerc et al. (2011) did not see a difference between ECs and non-elite controller slow progressors; whether these differences point to different roles of RICH2 in the different population groups remains to be elucidated.

Given that these RICH 2 SNPs, in combination appeared to be involved in disease control, we tried to determine their effect on RICH2 using a range of reputable online splice prediction tools to carry out extensive *in silico* analysis. When we analysed the combined effect of minor (mutant) alleles at the two SNP positions on the known acceptor splice site at intron10/exon 11 junction, although 3/5 tools predicted changes, it was mostly with respect to the 254^e SNP, rather than the combinations that revealed changes in scores. Analysis of the effect of the 255^i SNP indicated that a mutation at this site may introduce a new acceptor site or an alternative cryptic donor. The most consistent and convincing predictions however were with respect to the 254^e SNP that resides at the third nucleotide position within the first codon of exon 11, resulting in a synonymous amino acid change. A minor allele (G) at the 254^e SNP position was consistently predicted by all four analysis tools to result in the loss of an exonic splice enhancer (ESE) site (identified as SRSF2/SC35 by three tools). This result also verifies the FASTSNP prediction reported by Le Clerc et al. (2011). Although we cannot accurately predict what a loss of an ESE site will have on RICH2 expression, it is well known that ESEs promote exon splicing and are involved in alternative splicing (Lam and Hertel, 2002; Wang and Burge, 2008). If we postulate that a loss of the ESE (254^e minor allele) will result in altered RICH2 expression, given that a minor allele at the 254^e SNP is always associated with a minor allele at the 68 bp upstream 255^i SNP position in black sub-Saharan populations, and that this haplotype has almost become fixed in Caucasian populations (close to complete LD), a loss of an ESE site may necessitate a compensatory change to counteract or balance the altered RICH2 expression. The deleterious 254^eAA/255^iGA combination with respect to HIV-1 control and $CD4^+$ T-cell preservation in the presence of replicating virus, points to RICH2 expression, and consequently proinflammatory cytokine production, as a potentially important pathway in HIV-1 disease progression.

It has become increasingly clear that systemic chronic immune activation, rather than HIV-1 viral replication, is the main driving force of $CD4^+$ T-cell decline and the development of AIDS (reviewed (Paiardini and Muller-Trutwin, 2013)). It is not fully understood what

mechanisms contribute to chronic immune activation, and although still under debate, high levels of type I interferons and proinflammatory cytokines may contribute to increased activation (Paiardini and Muller-Trutwin, 2013). The associations seen with RICH2 and HIV-1 control imply that RICH2 mediated NF-κβ activation, through tetherin, is an important axis that needs to be investigated and may shed light on the role of the proinflammatory response in chronic immune activation and HIV-1 control. We hypothesise that increased RICH2 expression is likely to increase overall immune activation, and in a setting where viral replication is in itself contributing to increased immune activation, a mechanism that leads to the lower NF-κβ activation may be beneficial and aid in $CD4^+$ T-cell preservation and ultimately delay disease progression. It seems that HIV-1 needs to establish a fine balance with respect to NF-κβ activation since it is required for efficient LTR-dependent viral gene expression on the one hand and on the other hand induces immune factors that promote viral control (reviewed (Hotter et al., 2013)). A better understanding of the mechanisms that contribute to the regulation of NF-κβ activation may thus aid in finding novel ways of controlling harmful levels of chronic inflammation (Hotter et al., 2013).

A limitation of the current study is the small sample number within the groups being compared, although the study of extreme phenotypes serves to increase the power of a study (Barnett et al., 2013), these findings need to be tested using larger cohorts of sub-Saharan African viraemic controllers.

In conclusion, this study suggests a role for RICH2, and by proxy tetherin, in the natural control of HIV-1 in a black sub-Saharan population. *In vitro* studies that determine the role of the 254^{e}AA/255^{i}GA combined genotype on RICH2 expression will help determine if our proposed hypothesis of decreased immune activation as a consequence of the absence of this genotype in controllers with higher viraemia is plausible. In addition, a better understanding of the mechanisms and pathways that are involved in natural control of HIV-1 should ultimately aid in the development of novel therapeutics and vaccines.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is based on the research supported by grants awards from the Strategic Health Innovation Partnerships (SHIP) Unit of the South African Medical Research Council (a grantee of the Bill & Melinda Gates Foundation), the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation of South Africa and the National Institutes of Health, USA (R01HL090312 and P30AI094189: R. E. Chaisson).

References

Barnett IJ, Lee S, Lin X. Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. Genetic epidemiology. 2013; 37:142–151. [PubMed: 23184518]

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–265. [PubMed: 15297300]

Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annual review of genomics and human genetics. 2008; 9:403–433.

Casado C, Colombo S, Rauch A, Martinez R, Gunthard HF, Garcia S, Rodriguez C, Del Romero J, Telenti A, Lopez-Galindez C. Host and viral genetic correlates of clinical definitions of HIV-1 disease progression. PloS one. 2010; 5:e11079. [PubMed: 20552027]

Fellay J, Ge D, Shianna KV, Colombo S, Ledergerber B, Cirulli ET, Urban TJ, Zhang K, Gumbs CE, Smith JP, Castagna A, Cozzi-Lepri A, De Luca A, Easterbrook P, Gunthard HF, Mallal S, Mussini C, Dalmau J, Martinez-Picado J, Miro JM, Obel N, Wolinsky SM, Martinson JJ, Detels R, Margolick JB, Jacobson LP, Descombes P, Antonarakis SE, Beckmann JS, O'Brien SJ, Letvin NL, McMichael AJ, Haynes BF, Carrington M, Feng S, Telenti A, Goldstein DB. Immunology N.C.f.H.A.V. Common genetic variation and the control of HIV-1 in humans. PLoS genetics. 2009; 5:e1000791. [PubMed: 20041166]

Furuta B, Harada A, Kobayashi Y, Takeuchi K, Kobayashi T, Umeda M. Identification and functional characterization of nadrin variants, a novel family of GTPase activating protein for rho GTPases. Journal of neurochemistry. 2002; 82:1018–1028. [PubMed: 12358749]

Galao RP, Le Tortorec A, Pickering S, Kueck T, Neil SJ. Innate sensing of HIV-1 assembly by Tetherin induces NFkappaB-dependent proinflammatory responses. Cell host & microbe. 2012; 12:633–644. [PubMed: 23159053]

Galao RP, Pickering S, Curnock R, Neil SJ. Retroviral retention activates a Syk-dependent HemITAM in human tetherin. Cell host & microbe. 2014; 16:291–303. [PubMed: 25211072]

Gurdasani D, Iles L, Dillon DG, Young EH, Olson AD, Naranbhai V, Fidler S, Gkrania-Klotsas E, Post FA, Kellam P, Porter K, Sandhu MS. A systematic review of definitions of extreme phenotypes of HIV control and progression. Aids. 2014; 28:149–162. [PubMed: 24149086]

Hoffmann CJ, Cohn S, Mashabela F, Hoffmann JD, McIlleron H, Denti P, Haas DW, Dooley KE, Martinson NA, Chaisson RE. Treatment Failure, Drug Resistance, and CD4 T-Cell Count Decline Among Postpartum Women on Antiretroviral Therapy in South Africa. Journal of acquired immune deficiency syndromes. 2016; 71:31–37. [PubMed: 26334739]

Homann S, Smith D, Little S, Richman D, Guatelli J. Upregulation of BST-2/Tetherin by HIV infection in vivo. Journal of virology. 2011; 85:10659–10668. [PubMed: 21849457]

Hotter D, Sauter D, Kirchhoff F. Emerging role of the host restriction factor tetherin in viral immune sensing. Journal of molecular biology. 2013; 425:4956–4964. [PubMed: 24075872]

Hunt PW. Natural control of HIV-1 replication and long-term nonprogression: overlapping but distinct phenotypes. The Journal of infectious diseases. 2009; 200:1636–1638. [PubMed: 19852668]

Lam BJ, Hertel KJ. A general role for splicing enhancers in exon definition. Rna. 2002; 8:1233–1241. [PubMed: 12403462]

Le Clerc S, Coulonges C, Delaneau O, Van Manen D, Herbeck JT, Limou S, An P, Martinson JJ, Spadoni JL, Therwath A, Veldink JH, van den Berg LH, Taing L, Labib T, Mellak S, Montes M, Delfraissy JF, Schachter F, Winkler C, Froguel P, Mullins JI, Schuitemaker H, Zagury JF. Screening low-frequency SNPS from genome-wide association study reveals a new risk allele for progression to AIDS. Journal of acquired immune deficiency syndromes. 2011; 56:279–284. [PubMed: 21107268]

Limou S, Coulonges C, Herbeck JT, van Manen D, An P, Le Clerc S, Delaneau O, Diop G, Taing L, Montes M, van't Wout AB, Gottlieb GS, Therwath A, Rouzioux C, Delfraissy JF, Lelievre JD, Levy Y, Hercberg S, Dina C, Phair J, Donfield S, Goedert JJ, Buchbinder S, Estaquier J, Schachter F, Gut I, Froguel P, Mullins JI, Schuitemaker H, Winkler C, Zagury JF. Multiple-cohort genetic association study reveals CXCR6 as a new chemokine receptor involved in long-term nonprogression to AIDS. The Journal of infectious diseases. 2010; 202:908–915. [PubMed: 20704485]

Mahauad-Fernandez WD, Okeoma CM. The role of BST-2/Tetherin in host protection and disease manifestation. Immunity, inflammation and disease. 2016; 4:4–23.

Martinson NA, Gupte N, Msandiwa R, Moulton LH, Barnes GL, Ram M, Gray G, Hoffmann C, Chaisson RE. CD4 and viral load dynamics in antiretroviral-naive HIV-infected adults from Soweto, South Africa: a prospective cohort. PloS one. 2014; 9:e96369. [PubMed: 24831447]

McLaren PJ, Coulonges C, Bartha I, Lenz TL, Deutsch AJ, Bashirova A, Buchbinder S, Carrington MN, Cossarizza A, Dalmau J, De Luca A, Goedert JJ, Gurdasani D, Haas DW, Herbeck JT, Johnson EO, Kirk GD, Lambotte O, Luo M, Mallal S, van Manen D, Martinez-Picado J, Meyer L, Miro JM, Mullins JI, Obel N, Poli G, Sandhu MS, Schuitemaker H, Shea PR, Theodorou I, Walker BD, Weintrob AC, Winkler CA, Wolinsky SM, Raychaudhuri S, Goldstein DB, Telenti A, de Bakker PI, Zagury JF, Fellay J. Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112:14658–14663. [PubMed: 26553974]

Mellors JW, Kingsley LA, Rinaldo CR Jr, Todd JA, Hoo BS, Kokka RP, Gupta P. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. Annals of internal medicine. 1995; 122:573–579. [PubMed: 7887550]

Okulicz JF, Marconi VC, Landrum ML, Wegner S, Weintrob A, Ganesan A, Hale B, Crum-Cianflone N, Delmar J, Barthel V, Quinnan G, Agan BK, Dolan MJ. Infectious Disease Clinical Research Program HIVWG. Clinical outcomes of elite controllers, viremic controllers, and long-term nonprogressors in the US Department of Defense HIV natural history study. The Journal of infectious diseases. 2009; 200:1714–1723. [PubMed: 19852669]

Paiardini M, Muller-Trutwin M. HIV-associated chronic immune activation. Immunological reviews. 2013; 254:78–101. [PubMed: 23772616]

Paximadis M, Mathebula TY, Gentle NL, Vardas E, Colvin M, Gray CM, Tiemessen CT, Puren A. Human leukocyte antigen class I (A, B, C) and II (DRB1) diversity in the black and Caucasian South African population. Human immunology. 2012; 73:80–92. [PubMed: 22074999]

Paximadis M, Schramm DB, Gray GE, Sherman G, Coovadia A, Kuhn L, Tiemessen CT. Influence of intragenic CCL3 haplotypes and CCL3L copy number in HIV-1 infection in a sub-Saharan African population. Genes and immunity. 2013; 14:42–51. [PubMed: 23151487]

Peck J, Douglas Gt, Wu CH, Burbelo PD. Human RhoGAP domain-containing proteins: structure, function and evolutionary relationships. FEBS letters. 2002; 528:27–34. [PubMed: 12297274]

Ramsay M, Tiemessen CT, Choudhury A, Soodyall H. Africa: the next frontier for human disease gene discovery? Human molecular genetics. 2011; 20:R214–220. [PubMed: 21908518]

Rollason R, Korolchuk V, Hamilton C, Jepson M, Banting G. A CD317/tetherin-RICH2 complex plays a critical role in the organization of the subapical actin cytoskeleton in polarized epithelial cells. The Journal of cell biology. 2009; 184:721–736. [PubMed: 19273615]

Rotger M, Dalmau J, Rauch A, McLaren P, Bosinger SE, Martinez R, Sandler NG, Roque A, Liebner J, Battegay M, Bernasconi E, Descombes P, Erkizia I, Fellay J, Hirschel B, Miro JM, Palou E, Hoffmann M, Massanella M, Blanco J, Woods M, Gunthard HF, de Bakker P, Douek DC, Silvestri G, Martinez-Picado J, Telenti A. Comparative transcriptomics of extreme phenotypes of human HIV-1 infection and SIV infection in sooty mangabey and rhesus macaque. The Journal of clinical investigation. 2011; 121:2391–2400. [PubMed: 21555857]

Silvestri G, Paiardini M, Pandrea I, Lederman MM, Sodora DL. Understanding the benign nature of SIV infection in natural hosts. The Journal of clinical investigation. 2007; 117:3148–3154. [PubMed: 17975656]

The Genomes Project C. A global reference for human genetic variation. Nature. 2015; 526:68–74. [PubMed: 26432245]

Uitenbrook, DG. SISA-Binomial. 1997.

Walker BD. Elite control of HIV Infection: implications for vaccines and treatment. Topics in HIV medicine : a publication of the International AIDS Society, USA. 2007; 15:134–136.

Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. Rna. 2008; 14:802–813. [PubMed: 18369186]

WHO. WHO Media Centre Fact sheet. 2015. HIV/AIDS Fact sheet N°360.

Yuan HY, Chiou JJ, Tseng WH, Liu CH, Liu CK, Lin YJ, Wang HH, Yao A, Chen YT, Hsu CN. FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. Nucleic acids research. 2006; 34:W635–641. [PubMed: 16845089]

Highlights

- A combination genotype of two RICH2 SNPs is associated with natural HIV-1 control
- Genotype is deleterious with respect to control in presence of detectable viraemia
- *In silico* analysis predicts loss of an ESE site (SRSF2) with rs2072254 minor allele
- Progressing individuals with RICH2 rs2072255 (GA+AA) have lower $CD4^+$ T-cell counts
- Postulate altered RICH2 expression affecting tetherin's downstream NF-κβ activation

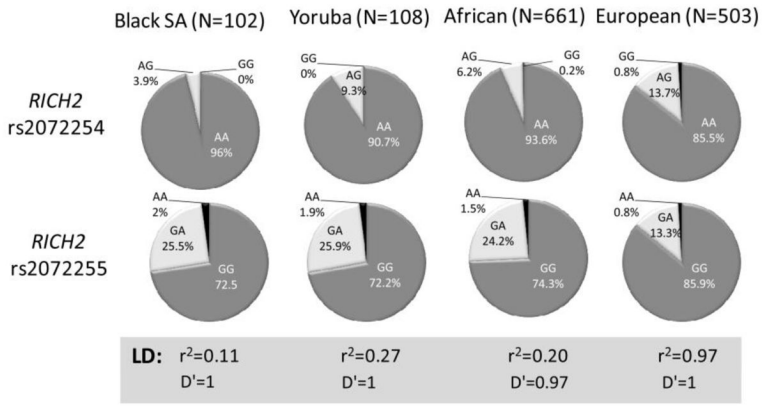

Figure 1.

Pie charts showing the genotypic frequencies of the exonic rs2072254 and intronic rs2072255 *RICH2* SNPs, as well as calculated linkage disequilibrium (LD), for the healthy black south African (SA) controls and three 1000 Genomes project reference populations [Nigerian Yoruba, total African and total European (Caucasian)].

A

Healthy Controls vs. HIV Controllers																			
Combined *RICH2* Genotypes		HCs vs. HICs			HCs vs. ECs			HCs vs. VCs			HCs vs. HVL LTNPs			HCs vs. HICs VL<400			HCs vs. HICs VL>400		
rs2072254	rs2072255	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
AA	GA	3.26	1.06-10.08	**0.04**	0.83	0.20-3.39	0.72	7.77	1.00-60.53	**0.02**	∞	NaN-∞	0.20	1.09	0.33-3.63	1.00	∞	NaN-∞	**0.002**
AG	GA	0.34	0.08-1.50	0.15	∞	NaN-∞	1.00	0.25	0.05-1.21	**0.09**	0.37	0.03-3.89	0.39	0.28	0.04-1.86	0.20	0.38	0.07 1.99	0.35
Progressors vs. HIV Controllers																			
Combined *RICH2* Genotypes		Progressors vs. HICs			Progressors vs. ECs			Progressors vs. VCs			Progressors vs. HVL LTNPs			Progressors vs. HICs VL<400			Progressors vs. HICs VL>400		
rs2072254	rs2072255	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
AA	GA	2.21	0.67-7.34	0.27	0.56	0.13-2.43	0.42	5.26	0.65-42.70	0.11	∞	NaN-∞	0.34	0.74	0.21-2.63	0.73	∞	NaN-∞	**0.02**
AG	GA	0.59	0.15-2.33	0.50	∞	NaN-∞	1.00	0.44	0.10-1.90	0.27	0.63	0.06-6.31	0.54	0.49	0.08- 2.96	0.60	0.66	0.14-3.13	0.68

B

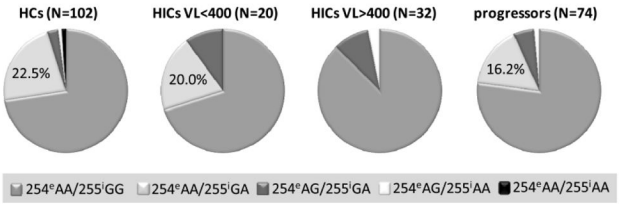

Figure 2.

(**A**) Box with comparisons of predominant genotype combinations of the two *RICH2* SNPs (excluding homozygosity for major alleles for both SNPs) between healthy controls (HCs) and HIV-1 infected controllers (HICs) and controller subgroups (ECs: elite controllers, VCs: viraemic controllers, HVL LTNPs: high viral load long term non-progressors, progressors, HICs VL<400: controllers stratified according to viral load less than 400 RNA copies/ml, HICs VL>400: controllers stratified according to viral load greater than 400 RNA copies/ml) and between HIV-1-infected progressors and HIV-1 infected controllers (HICs) and controller subgroups (as above). Odds Ratios (OR), confidence intervals (CI) and *P* values are listed (*P*s<0.05 are shown in bold). (**B**) Pie charts showing representation (% frequency) of all detected genotype combinations in HCs, HICs VL<400, HICs VL>400 and HICs VL>400 and progressors. 254^{e} denotes the rs2072254 exonic SNP and 255^{i} denotes the rs2072255 intronic SNP. Only the frequency value (%) of the significantly different genotypic combination (254^{e}AA/255^{i}GA) is shown on respective charts.

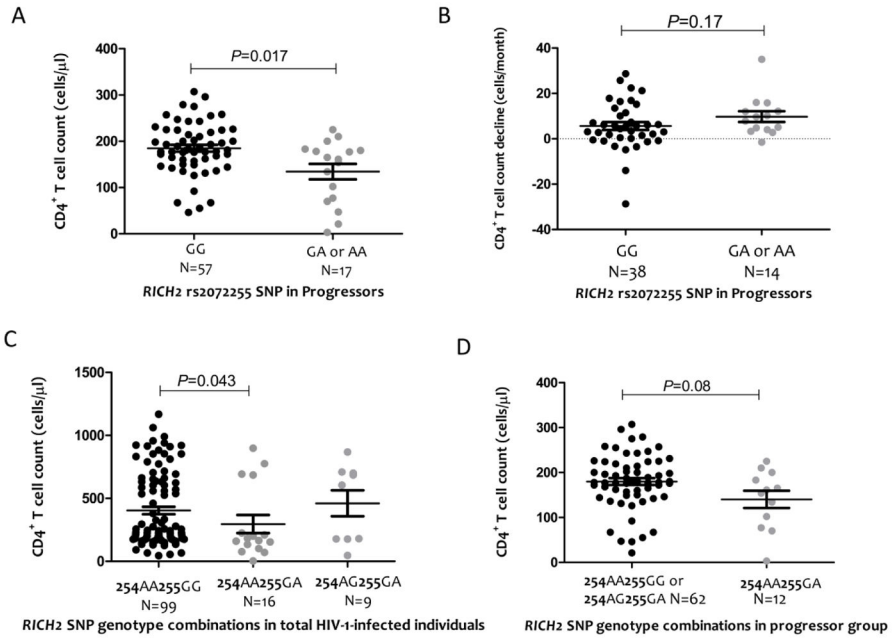

Figure 3.

Graphs showing comparisons of (**A**, **C** and **D**) $CD4^+$ T-cell counts (cells/μl) and (**B**) rate of $CD4^+$ T-cell decline (cells/month) between individuals with select genotypes and in respective study groups as shown below each graph. 254 and 255 in graphs C and D denote the rs2072254 and rs2072255 *RICH2* SNPs, respectively. Bold solid long horizontal line depicts median with 25th and 75th percentiles (shorter horizontal lines) shown. *P* values are shown between comparisons as indicated.

Table 1

Charcteristics of HIV-1-infected black South African study groups classified according to varying levels of control

Clinical Phenotype Groups	N	Age (years) Mean and Range	Gender % female	CD4+ T cell counts[1] (cells/μl blood) Median and IQRs[3]	Viral load[1] (HIV-1 RNA copies/ml plasma) Median and IQRs	Time since HIV-1 diagnosis (years)[2] Median and IQRs
Combined HIV Controllers (HICs)	52	37 (19 – 54)	86.5	697 (604 – 879)	1123 (127 – 2076)	8 (2–11)
HIC sub-groups						
Elite Controllers **(ECs)**	11	44 (19 – 54)	81.8	853 (718 – 1022)	<40 (<40 (n=2), <20 (n=9))	11 (9–12)
Viraemic Controllers **(VCs)**	30	36 (19 – 46)	90.0	651 (555 – 832)[4]	495 (327 – 965)[5]	3 (0–10.35)
High Viral Load long term non-progressors **(HVL LTNPs)**	11	41 (31 – 51)	81.8	663 (635 – 749)	54 375 (13 415 – 77 820)	8 (8–11)
HIV Controllers (HICs) with viral loads < 400 HIV-1 RNA copies/ml **(HICs VL<400)**	20	39 (19–54)	85.0	693 (592 – 907)	138 (<20 – 155)	10 (2–12)
HIV Controllers (HICs) with viral loads > 400 HIV-1 RNA copies/ml **(HICs VL>400)**	32	36 (19 – 51)	87.5	707 (604–865)	1610 (917 – 15 709)	7 (2.25–9.75)
Progressors	74	43 (28 – 71)	83.8	177 (146 – 210)	38 444 (19 853 – 103 042)	6 (1–7)

[1] $CD4^+$ T-cell counts and viral loads of HIV Controllers reported are from study enrolment time point, whereas for progressors the $CD4^+$ T-cell counts and viral loads are from the last clinic visit prior to initiation of treatment.

[2] This period was calculated from the year of HIV-1 diagnosis to the last year of monitoring that included a $CD4^+$ T-cell count and viral load determination.

[3] Interquartile ranges.

[4] Three individuals had CD4+ T cell counts <500 cells/μl at time of enrolment; however were included based on length of time of infection since diagnosis to time of enrolment, in the absence of ART (≥9 years).

[5] One individual had a VL>2 000 (VL=6070), but was classified as a VC also due to length of infection since diagnosis with viral control in the absence of ART (9 years).

Table 2

Comparison of allelic and genotypic frequencies of the two *RICH2* SNPs between healthy controls and HIV-1-infected controllers and controller subgroups as well as between Progressors and HIV-1-infected controllers and controller subgroups

Healthy Controls vs.HIV-1 Controllers																		
***RICH2* SNP rs2072254**	**HCs vs. Combined HICs**			**HCs vs. ECs**			**HCs vs. VCs**			**HCs vs. HVL LTNPs**			**HCs vs. HICs VL<400**			**HCs vs. HICs VL>400**		
Allele/Genotype	**OR**	**CI**	***P***	**OR**	**CI**	***P***	**OR**	**CI**	***P***	**OR**	**CI**	***P***	**OR**	**CI**	***P***	**OR**	**CI**	***P***
G	0.33	0.09–1.18	**0.09**	∞	NaN-∞	1.00	0.28	0.07–1.16	**0.08**	0.20	0.03–1.16	0.11	0.38	0.07–2.15	0.26	0.30	0.07–1.24	0.10
AG	0.31	0.08–1.16	**0.09**	∞	NaN-∞	1.00	0.27	0.06–1.13	**0.08**	0.18	0.03–1.14	0.10	0.37	0.06–2.16	0.25	0.29	0.07–1.22	**0.09**
***RICH2* SNP rs2072255**	**HCs vs. Combined HICs**			**HCs vs. ECs**			**HCs vs. VCs**			**HCs vs. HVL LTNPs**			**HCs vs. HICs VL<400**			**HCs vs. HICs VL>400**		
Allele/Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
A	1.46	0.70–3.04	0.38	1.09	0.30–3.92	1.00	1.90	0.70–5.13	0.28	1.09	0.30–3.92	1.00	0.98	0.38–2.53	1.00	2.03	0.76–5.49	0.20
GA	1.64	0.70–3.83	0.31	0.94	0.23–3.80	1.00	1.76	0.61–5.07	0.34	3.16	0.38–26.18	0.45	0.82	0.29–2.36	0.78	3.28	0.92–11.70	**0.08**
AA	1.4	0.1–2.90	1.00	∞	NaN-∞	1.00	∞	NaN-∞	1.00	0.24	0.02–2.96	0.31	∞	NaN-∞	1.00	0.76	0.07–8.68	1.00
GA+AA	1.59	0.70–3.59	0.32	1.01	0.25–4.08	1.00	1.89	0.66–5.43	0.34	1.70	0.35–8.37	0.72	0.88	0.31–2.52	0.79	2.65	0.85–8.24	0.10
Progressors vs. HIV-1 Controllers																		
***RICH2* SNP rs2072254**	**Progressors vs. Combined HICs**			**Progressors vs. ECs**			**Progressors vs. VCs**			**Progressors vs. HVL LTNPs**			Progressors vs. HICs VL<400			**Progressors vs. HICs VL>400**		
Allele/Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	***P***	OR	CI	*P*
G	0.57	0.17–1.92	0.37	∞	NaN-∞	1.00	0.49	0.13–1.89	0.28	0.35	0.06–1.92	0.22	0.66	0.12–3.56	0.64	0.52	0.14–2.02	0.46
AG	0.56	0.16–1.93	0.36	∞	NaN-∞	1.00	0.47	0.12–1.89	0.28	0.33	0.06–1.94	0.22	0.65	0.12–3.64	0.64	0.51	0.13–2.03	0.45
***RICH2* SNP rs2072255**	**Progressors vs. Combined HICs**			**Progressors vs. ECs**			**Progressors vs. VCs**			**Progressors vs. HVL LTNPs**			**Progressors vs. HICs VL<400**			**Progressors vs. HICs VL>400**		
Allele/Genotype	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*	OR	CI	*P*
A	1.17	0.53–2.60	0.84	0.88	0.24–3.26	0.74	1.52	0.54–4.31	0.63	0.88	0.24–3.26	0.74	0.29	0.29–2.13	0.60	1.63	0.58–4.61	0.47
GA	1.31	0.53–3.25	0.65	0.75	0.18–3.15	0.71	1.40	0.46–4.25	0.60	2.53	0.30–21.45	0.68	0.66	0.22–1.98	0.55	2.62	0.70–9.74	0.17
AA	0.74	0.05–12.12	1.00	∞	NaN-∞	1.00	∞	NaN-∞	1.00	0.16	0.01–2.76	0.27	∞	NaN-∞	1.00	0.49	0.03–8.15	1.00

Healthy Controls vs.HIV-1 Controllers																		
***RICH2* SNP rs2072254**	**HCs vs. Combined HICs**			**HCs vs. ECs**			**HCs vs. VCs**			**HCs vs. HVL LTNPs**			**HCs vs. HICs VL<400**			**HCs vs. HICs VL>400**		
Allele/Genotype	**OR**	**CI**	***P***	**OR**	**CI**	***P***	**OR**	**CI**	***P***	**OR**	**CI**	***P***	**OR**	**CI**	***P***	**OR**	**CI**	***P***
GA+AA	1.25	0.52–3.01	0.66	0.79	0.19–3.33	0.72	1.49	0.50–4.49	0.60	1.34	0.26–6.82	1.00	0.70	0.23–2.09	0.56	2.09	0.64–6.79	0.29

HC: Healthy HIV-1 negative controls; **HICs:** HIV-1-infected controllers; **ECs:** Elite Controllers; **VCs**: Viraemic Controllers; **HVL LTNPs:** High Viral Load long term non-progressors; **HICs VL<400:** HIV-1-infected controllers with viral loads less than 400 HIV-1 RNA copies/ml plasma; **HICs VL>400:** HIV-1-infected controllers with viral loads greater than 400 HIV-1 RNA copies/ml plasma.

Significant *P* values (<0.05) and *P* values showing strong trends (0.05>*P*<0.1) are shown in bold.

Table 3

Number and frequency (%) of the combined genotypes of the two *RICH2* SNPs in HIV-1-uninfected healthy controls, HIV-1- infected controllers and controller sub-groups and HIV-1-infected Progressors

Combined	Genotypes	clinical phenotype groups			HICs sub-groups			HICs sub-groups	
		HCs N=102	Combined HICs N=52	Progressors N=74	ECs N=11	VCs N=30	HVL LTNPs N=11	HICs VL<400 N=20	HICs VL>400 N=32
rs2072254	**rs2072255**								
AA	**GG**	74 (72.55)	42 (80.77)	57 (77.03)	8 (72.73)	25 (83.33)	9 (81.82)	14 (70)	28 (87.50)
AA	**GA**	23 (22.55)	4 (7.69)	12 (16.22)	3 (27.27)	1 (3.33)	0 (0.00)	4 (20)	0 (0.00)
AG	**GA**	3 (2.94)	5 (9.62)	4 (5.41)	0 (0.00)	4 (13.33)	1 (9.09)	2 (10)	3 (9.38)
AG	**AA**	1 (0.98)	1 (1.92)	1 (1.35)	0 (0.00)	0 (0.00)	1 (9.09)	0 (0.00)	1 (3.13)
AA	**AA**	1 (0.98)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)

HC: Healthy HIV-1 negative controls; **HICs:** HIV-1-infected controllers; **ECs:** Elite Controllers; **VCs**: Viraemic controllers; **HVL LTNPs:** High Viral Load long term non-progressors; **HICs VL<400:** HIV-1-infected controllers with viral loads less than 400 HIV-1 RNA copies/ml plasma; **HICs VL>400:** HIV-1-infected controllers with viral loads greater than 400 HIV-1 RNA copies/ml plasma.

Table 4

In silico analysis of two *RICH2* SNPs and their potential impact on *RICH2* splicing

Software Tool	**Scores of known acceptor splice site (ss) at intron 10 and exon 11 junction**				
	Combined *RICH2* SNP alleles				
	255^i WT/254^e WT[1]	**255^i WT/254^e mt**	**255^i mt/254^e WT**	**255^i mt/254^e mt**	**Result**
SplicePort[2]	2.12	2.09	2.08	2.04	↓both SNPs
ASSP	5.46	5.46	5.46	5.46	No change
BDGP	0.79	0.77	0.79	0.77	↓254^e SNP
ESEfinder	6.15	6.34	6.15	6.34	↑254^e SNP
HSF	81.3	81.3	81.3	81.3	No change
	254^e SNP alleles				
	WT		mt		Result
MaxEntScan[3]	6.80		6.53		↓254^e SNP
Software Tool	Altered splice site predictions for *RICH2* rs2072255 intronic SNP				
SplicePort[4]	Introduction of new acceptor ss (ttgcgaggtctc) with mt allele: score=−4.2				
ASSP	Introduction of alternative isoform/cryptic donor ss (TGTCCTGTTGgtagggactc) with mt allele: score=5.47				
HSF	Introduction of new acceptor site (ggcagttgcgagGT) with mt allele: score=71.72				
Software Tool	Analysis of *RICH2* rs2072254 exonic SNP				
ESEfinder	SRSF2 binding site (ggactctt) lost with mt allele: score with WT allele=2.42				
AASsites	ESE analysis predicted loss of SC35 binding site with mt allele (no score provided)				
HSF	SC35 site is reported to be abolished with mt allele: score with WT allele=75.2				
SKIPPY	Loss of strongly predicted ESE site with mt allele (no score provided)				

^{i} intronic;

^{e} exonic;

WT=major allele; mt=minor allele; 254: *RICH2* rs2072254 SNP; 255: *RICH2* rs2072255 SNP; ESE: exon splicing enhancer; ↓: decrease in score; ↑: increase in score

[1] Since the intron 10/exon 11 junction is known for *RICH2*, all the software predicted acceptor splice site (ss) with high scores for this junction as expected. These scores (i.e. the 255^{i}WT/254^{e}WT scores) were thus used to investigate the effects of the two *RICH2* SNP alleles in the respective combinations on the known acceptor splice.

[2] SpicePort score threshold set to '0': 87% sensitivity and 1.4% false positive rate (FPR).

[3] MaxEntScan calculates a maximum entropy score and allows input of only 23 nucleotides over the intron/exon boundry (AG site) for 3′ ss prediction and thus could not assess effect of the intronic rs2072255 SNP.

[4] SplicePort default score threshold was used (−10): 100% sensitivity and 85.2% FPR for AG locations (acceptor ss) and 86.9% FPR for locations (donor ss), i.e. less stringent than [2] above.