Role of HIV-specific CD8+ T cells in pediatric HIV cure strategies after widespread early viral escape

Ellen M. Leitman,1* Christina F. Thobakgale,4* Emily Adland,1* M. Azim Ansari,2 Jayna Raghwani,3 Andrew J. Prendergast,5,6 Gareth Tudor-Williams,7 Photini Kiepiela,8,9 Joris Hemelaar,10,12 Jacqui Brener,1 Ming-Han Tsai,1 Masahiko Mori,1,13 Lynn Riddell,14 Graz Luzzi,15 Pieter Jooste,16 Thumbi Ndung’u,4,17 Bruce D. Walker,4,17 Oliver G. Pybus,3 Paul Kellam,18,19 Vivek Naranbhai,17,20 Andrew J. Prendergast,5,6 Gareth Tudor-Williams,7 Photini Kiepiela,8,9 Joris Hemelaar,10,12

INTRODUCTION

Recent studies have suggested greater HIV cure potential among infected children than adults. A major obstacle to HIV eradication in adults is that the viral reservoir is largely comprised of HIV-specific cytotoxic T lymphocyte (CTL) escape variants. We here evaluate the potential for CTL in HIV-infected slow-progressor children to play an effective role in "shock-and-kill" cure strategies. Two distinct subgroups of children were identified on the basis of viral load. Unexpectedly, in both groups, as in adults, HIV-specific CTL drove the selection of escape variants across a range of epitopes within the first weeks of infection. However, in HIV-infected children, but not adults, de novo autologous variant-specific CTL responses were generated, enabling the pediatric immune system to "corner" the virus. Thus, even when escape variants are selected in early infection, the capacity in children to generate variant-specific anti-HIV CTL responses maintains the potential for CTL to contribute to effective shock-and-kill cure strategies in pediatric HIV infection.

*E.M. Leitman, C.F. Thobakgale, and E. Adland contributed equally to this paper.

© 2017 Leitman et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike 4.0 International license, as described at https://creativecommons.org/licenses/by-nc-sa/4.0/.

© 2017 Leitman et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike 4.0 International license, as described at https://creativecommons.org/licenses/by-nc-sa/4.0/.

https://doi.org/10.1084/jem.20162123
slow progressors (PSPs) typically have viral loads of \(\sim 30,000\) HIV copies/ml and generally do not express HLA class I molecules that are protective against adult disease progression (Adland et al., 2015; Muenchhoff et al., 2016). These children thus share some features of nonpathogenic SIV infection observed in natural nonhuman primate hosts such as the sooty mangabey, that is characterized by preservation of normal peripheral CD4 counts in the face of persistently high viral loads (Rey-Cuillé et al., 1998). In addition, the localization of the virus predominantly in the short-lived memory CD4\(^+\) T cells is a second aspect of immunodeficiency virus infection that is shared by HIV-infected pediatric nonprogressors and SIV-infected sooty mangabeys (Paiardini et al., 2011; Muenchhoff et al., 2016). In combination with the tolerogenic immune environment in early life that limits the establishment of the viral reservoir, these features together suggest a greater potential for HIV cure in children than adults (Goulder et al., 2016).

To further investigate HIV cure potential among PSPs, defined here as perinatally infected children who did not meet prevailing clinical (World Health Organization [WHO] stage III/IV clinical disease) or CD4 (maintained CD4 >350 cells/mm\(^3\)) criteria for ART initiation by the age of 5 yr, we studied HIV-specific CD8\(^+\) T cell activity among these children, because, according to “shock-and-kill” strategies proposed (Deeks, 2012), these cells would play a critical role in clearance of HIV-infected cells after latency reversal in ART-treated in (Rey-Cuillé et al., 2015). In contrast, although HIV-specific CD8\(^+\) T cells are the major part of the adult viral reservoir (Deng et al., 2015), in that multiple escape mutants are rapidly selected early in the course of adult infection, and these escape viruses constitute the major part of the adult viral reservoir (Deng et al., 2015). In contrast, although HIV-specific CD8\(^+\) T cell responses are detectable early in life, viral setpoint is not established in infected children until 5 yr of age, even in those maintaining normal-for-age CD4 counts (Muenchhoff et al., 2016). Furthermore, studies of early pediatric progression have failed to demonstrate the selection of escape mutants, even when there are detectable CTL responses restricted by HLA alleles that are “protective” in adult infection, such as HLA-B*57/58:01/81:01 (Adland et al., 2015). This prompts the hypothesis that HIV-specific CD8\(^+\) T cells do not impose significant selection pressure on HIV in early life sufficient for the widespread selection of escape mutants and, therefore, that viral reservoirs in infected children on ART are populated largely by wild-type virus. However, studies of CD8\(^+\) T cell escape have not been undertaken to date in PSPs, with access to maternal samples and longitudinal pediatric samples in combination with next-generation sequencing, that together allow this issue to be interrogated.

To examine the impact of HIV-specific CD8\(^+\) T cell responses on circulating virus in PSPs, we determined the HIV-specific CD8\(^+\) T cell responses and ultra-deep sequenced virus within longitudinal samples from 11 PSPs from a historical cohort of perinatally infected children and their mothers followed from birth through the first decade of life, recruited in 2003–2005 to the Pediatric Early HAART and Strategic Treatment Interruption Study (Mphatswe et al., 2007; Prendergast et al., 2008). We hypothesized, based on previous studies in adults (Kiepiela et al., 2007), that the breadth, specificity and function, reflected by the selection of virus escape mutants, of HIV-specific CD8\(^+\) T cell responses would differ according to viral load in the PSPs. We report here that two distinct HIV-specific CD8\(^+\) T cell immune strategies can be discerned among PSPs: one is similar to that adopted by adult elite controllers and viremic controllers (VCs), and the other is reminiscent of that described among adult viremic non-controllers (VNCs; Kiepiela et al., 2007).

**RESULTS**

**PSPs categorized into high- and low-viremic groups**

Of 75 perinatally infected South African infants followed longitudinally from birth (Mphatswe et al., 2007; Prendergast et al., 2008), before universal ART recommendation for pediatric HIV infection, 11 children had failed to meet prevailing clinical (WHO stage III/IV clinical disease) or CD4 criteria (absolute CD4 count of <350 cells/mm\(^3\)) for ART initiation by 5 yr of age. These children we define here as PSPs (Table 1).

Hypothesizing that, as in adults, the HIV-specific CD8\(^+\) T cell responses might differ with respect to specificity, breadth, and function, PSPs were subdivided into two subgroups according to the median viral load: one with persistently higher and one with persistently lower viral loads (Fig. 1 A). These we have termed VCs and VNCs. Although the viral loads overall were \(0.87 \log_{10}\) lower in the VNC group (\(P = 5 \times 10^{-14}\)), there was no significant difference between the groups in either absolute CD4 count or CD4\% (Fig. 1, B and C), which remained healthy (median 735 cells/mm\(^3\) and 29.5\%) and very similar in both groups, after the natural CD4 and CD4\% decline over the first 5 yr that is observed in healthy uninfected children (Shearer et al., 2003; Lugada et al., 2004).

**Broad, high-magnitude non–Gag-specific CD8 responses among VNC PSPs in early life**

Specificity and breadth of the HIV-specific CD8\(^+\) T cell response has been related to viral load in previous studies of adult HIV infection, with broad Gag-specific responses associated with lower viral loads, and broad Env-, Nef- and accessory-regulatory protein (Vif/Vpr/Vpu/Tat/Rev)-specific responses associated with higher viral loads (Klein et al., 1995; Rivière et al., 1995; Ogg et al., 1998; Barouch et al., 2002, 2003; Edwards et al., 2002; Novitsky et al., 2003; Masemola et al., 2004; Zuniga et al., 2006; Geldmacher et al., 2007; Kiepiela et al., 2007; Streeck et al., 2007; Janes et al., 2013). We therefore next investigated whether differential CD8\(^+\) T cell activity might explain the distinct viral loads among the two groups of PSPs. Using a panel of 410 overlapping 18-mer peptides spanning the clade C proteome, we determined in IFN-\(\gamma\) ELISPOT assays which individual peptides
were recognized (Addo et al., 2003; Kiepiela et al., 2004). In both groups, HIV-specific CD8+ T cell responses significantly increased with age in breadth and magnitude (Fig. 2, A and B). However, total breadth was 1.5 responses higher in VNC than in VC PSPs ($P = 3 \times 10^{-9}$; Fig. 2 A). This difference in breadth was most significant in Nef-, Acc-Reg-, and Env-specific responses ($P = 4 \times 10^{-10}, 4 \times 10^{-5}, \text{and} \ 4 \times 10^{-4}$, respectively), but not in Gag ($P = 0.07$; Fig. 2 A). A similar pattern was observed for the magnitude of CD8+ T cell responses (Fig. 2 B). Overall, breadth and magnitude of non–Gag/Pol-specific responses were higher than breadth and magnitude of Gag/Pol-specific responses in the VNC children, whereas the reverse pattern was observed in the VC children (Fig. 3). These pediatric data are therefore consistent with adult studies indicating increased viral loads associated with non–Gag-specific responses, particularly those directed toward Env or Nef.

Selection pressure in VC and VNC children targeting Gag-TL9

To investigate the impact on the virus of HIV-specific CD8+ T cells in these two distinct groups of PSPs, we next tracked the HIV-specific CD8+ T cell responses and kinetics of selection pressure imposed on HIV from birth through childhood in each child also having access to samples from the transmitting mother. To maximize the resolution of adaptive sequence changes, we deep sequenced the virus. Phylogenetic analysis of full-length consensus sequences for all 11 mother–child pairs confirmed the close relationship between virus in each mother–child pair (Fig. S1). In each child, we analyzed the regions of the virus encoding the CD8+ T cell epitopes restricted by any of the six HLA alleles expressed by that child using a panel of well-characterized epitopes (Goulder et al., 1997, 2000, 2001a; Honeyborne et al., 2006; Matthews et al., 2008, 2011, 2012; Ngumbela et al., 2008; Kløverpris et al., 2012a,b, 2013, 2014a,b; Llano et al., 2013) representing all the HIV proteins (see Materials and methods). For this analysis, variants shared with the mother expressing the same HLA class I molecule as the child were presumed to have been transmitted and not to have arisen de novo in the child. If the variant arose after the initial time point in the child, it was assumed that this variant was selected in the child and was not transmitted.

We first focused on the specific CD8+ T cell responses and viral sequence changes in three children, all of whom targeted the same immunodominant Gag-TL9 epitope (TPQDLNTML, Gag residues 180–188). These three pediat-

| Table 1. Clinical characteristics of the PSPs |
|---|---|---|---|---|---|---|---|---|
| PID | DOB | Transmission | Sex | C or M | HLA-A | HLA-B | HLA-C | Early ART duration | Early ART start (age) | Early ART stop (age) | ART (re)-start age | Type of slow progression |
| 133-C | 11/2003 | IU | f | C | 68:01 | 74:01 | 58:02 | 81:01 | 04:01 | 06:02 | yr | none | yr | n/a | yr | n/a | left study at 9 yr, no ART | VC group |
| 468-C | 12/2004 | IU | m | C | 03:02 | 68:01 | 57:03 | 58:01 | 06:02 | 07:01 | 1.5 | 0.1 | 1.6 | 6.1 |
| 349-C | 11/2004 | IU | f | C | 23:01 | 43:01 | 45:01 | 58:02 | 06:02 | 07:01 | 1.5 | 0.1 | 1.6 | 6.1 |
| 364-C | 09/2004 | IU | f | C | 29:02 | 30:02 | 41:02 | 44:03 | 07:01 | 07:01 | 1.5 | 0.1 | 1.6 | 6.1 |
| 385-C | 01/2005 | IP | f | C | 30:01 | 34:02 | 15:03 | 58:02 | 02:10 | 06:02 | 1.3 | 0.1 | 1.6 | 6.1 |
| 517-C | 01/2005 | IP | m | C | 30:02 | 43:01 | 15:03 | 58:02 | 02:02 | 06:02 | 1.3 | 0.1 | 1.6 | 6.1 |
| 021-C | 09/2005 | IP | f | C | 02:05 | 29:02 | 42:01 | 44:03 | 02:10 | 07:01 | 1.0 | 0.1 | 1.1 | never |
| 114-C | 09/2003 | IP | f | C | 03:01 | 24:02 | 07:02 | 08:01 | 07:01 | 07:02 | 1.0 | 0.1 | 1.1 | 7.3 yr, no ART |
| 559-C | 02/2005 | IP | m | C | 23:01 | 68:02 | 08:01 | 57:02 | 07:01 | 07:01 | 1.0 | 0.1 | 1.1 | 7.3 yr, no ART |
| 380-C | 10/2004 | IP | m | C | 23:01 | 68:02 | 15:10 | 15:10 | 03:04 | 07:01 | 1.0 | 0.1 | 1.1 | 7.3 yr, no ART |
| 081-C | 08/2003 | IP | f | C | 29:02 | 68:02 | 15:16 | 44:03 | 03:04 | 07:01 | 1.4 | 0.1 | 1.5 | 6.7 |

For each subject, child’s HLA type is shown in the first row and maternal HLA is in the second row. C, child; DOB, date of birth; f, female; IP, intrapartum infection; IU, in utero infection; m, male; M, mother; n/a, not applicable; PID, participant identification.
ric subjects comprised one VC child, 133-C, and two VNC children, 517-C and 021-C.

The VC slow progressor 133-C made a dominant HLA-B*81:01–restricted response from the first time point tested (Figs. 4, A–C). The virus transmitted to this child encoded the wild-type TL9 epitope, and a T186M variant was selected that had almost reached fixation (96% frequency) by 2.7 yr of age (Fig. 4 D). Population sequencing showed that the T186M mutant arose between 0.7 and 1.3 yr (Thobakgale et al., 2009). A T186M variant–specific CD8+ T cell response was detectable by 1.3 yr of age and ultimately dominated the TL9-specific response (Fig. 4 E). The more common
In both VNC slow progressors, 517-C and 021-C, the Gag-TL9 responses were restricted by HLA-B*81:01 as opposed to HLA-B*81:01, arose later in the course of infection (Fig. 5), and did not become immunodominant for several years. In 517-C, the dominant response initially was to the HLA-C*03-restricted Gag-YL9 epitope (YVDRFFKTL, Gag 286–304). Early Gag escape mutants were not observed in 517-C, and no Gag-TL9 escape mutants were observed at any time point studied through 8.5 yr (Fig. 5 D and Tables 3, S1, and S2). However, a broad array of non-Gag escape mutants were selected to fixation within 2.4 mo of birth (517-C was infected intrapartum and only became viremic at 28 d), for example within the HLA-B*15:10-restricted Rev epitope I9 (IHSISERIL, Rev residues 52–60; Table 3). HLA-B*15:10 was not expressed in the mother 517-M and, as expected, the maternal virus transmitted encoded the wild-type Rev-I9 epitope. Other epitopes in which escape to fixation appeared to have occurred by 2.4 mo of age in 517-C were in two HLA-B*42:01-restricted Env epitopes, R110 (RPNNNT RKSI, Env residues 298–307) and IF9 (IPRRIRQGF; Env residues 843–851), and in two HLA-A*30:02-restricted Env epitopes, KY9 (KYLGSIVQY, Env residues 794–802) and IY9 (IVNRVQRQY, Env residues 704–712). In all four cases, the epitope mutant in the child's viral population differed in 100% of viral sequences from the maternal sequence. The caveat here would be that both mother and child express HLA-A*30:02 and HLA-B*42:01 and that the samples sequenced were of plasma virus 2–3 mo after transmission. However, particularly in view of the clear-cut evidence of early escape within the HLA-B*15:10 Rev I9 epitope, it would seem very likely that early escape is driven to fixation in these four Env epitopes, with additional evidence of early escape in almost 50% of sequences within the HLA-A*30:02-restricted Pol epitope AY9 (AQNPQIVY, Pol residues 328–366) and in 60% of sequences within the HLA-B*15:10-restricted Vif epitope WI9 (WHLGHGVSIF, Vif residues 79–83; Table 3).

In the other VNC child, 021-C, escape in the Gag-TL9 epitope was observed, but only relatively late, with no epitope variants being evident at 2.6 yr (Fig. 5 H). Again, early escape within 2 mo of birth was evident in non-Gag epitopes. HLA-A*29:02 was not expressed in the mother (Table S2). This low-frequency variant was ultimately replaced by the characteristic HLA-A*29:02 escape footprint in this epitope [Carlson et al., 2012]. Another instance of early escape within 2 mo of birth was evident again at low frequency within the HLA-A*02-restricted Pol-VL9 (VIYQYMDDDL, Pol residues 334–342). HLA-A*02 was not shared with the mother (Table S2). However, this variant did not persist and reverted back to wild-type in 100% sequences. There were no other instances of early escape identified in this child's viral sequences, although many escape mutants were transmitted, and in this mother–child pair, five of the six HLA class I molecules were shared.

Thus, in the VNC child 133-C, there was evidence of strong selection pressure within the first 3 mo of life, with broad Gag-specific CD8+ T cell responses generated (to HLA-A*02–Gag-SK10 and HLA-B*81–Gag-TL9) and escape mutants within Gag, Pol, and Vif, all reaching fixation.
in that time. In contrast, although escape mutants were selected within the same time period in the two VNC children (517-C and 021-C), these were mostly in Env, and none were in Gag. These data are consistent with the notion of control of broad Gag-specific CD8+ T cell responses contributing to lower viremia in early pediatric infection, as previously shown in adult infection (Klein et al., 1995; Riviére et al., 1995; Ogg et al., 1998; Edwards et al., 2002; Novitsky et al., 2003; Masemola et al., 2004; Zuñiga et al., 2006; Geldmacher et al., 2007; Kiepiela et al., 2007; Streeck et al., 2007).

**Preferential recognition of autologous Gag–TL9 variant in children, but not adults**

The observation that preferential recognition of the autologous variant is more common in HIV-infected children than adults was previously made with respect to the HLA-B*57/58:01–restricted Gag–TW10 epitope (Feeney et al., 2005). To investigate whether this might apply more generally, we here investigated whether preferential recognition of the autologous Gag–TL9 variant is also observed more commonly in pediatric than adult infection. We identified 11
children and 16 adults from Southern Africa infected with C clade virus who expressed either HLA-B*42:01 and/or HLA-B*81:01 and in whom autologous virus encoded a Gag-TL9 variant (Table S3). In two of the HIV-infected children, population sequencing of autologous virus initially demonstrated the presence of a mixture of variants at Gag-TL9 variants at Gag-182 (position 3 in the epitope; Fig. 6, A and B), suggesting recent selection of escape (Goonetilleke et al., 2009). Preferential recognition of the single autologous Gag-TL9 variant that subsequently emerged was demonstrated in both of these two children (Fig. 6, A and B). Overall, in the 11 children studied, preferential recognition of the autologous Gag-TL9 variant was observed in all but one (91%), whereas in the 16 adults studied, preferential recognition of the autologous variant was observed in only 5 (31%; P = 0.0047; Fig. 6, C and D).

The caveats here (discussed further below) are that the transmitted virus in all of the adults and some of the children were unknown. However, in 3 of the children, an evolving preferential recognition of the autologous Gag-TL9 variant was demonstrated (Figs. 4 and 6), and in 4 more children, maternal virus encoded wild-type Gag-TL9 and/or the mother was HLA-B*42:01/81:01 negative. Even if one restricts the analysis to these 7 children, the observed preferential recognition of autologous Gag-TL9 variants remains significantly greater in children than adults (P = 0.0045, Fisher’s exact test).

Contrasting selection pressure in VC and VNC children
To examine selection pressure more broadly and compare the differences between VNC and VC children in the kinetics of accumulation of escape mutants, we calculated the proportion of epitopes within the most immunogenic proteins (Gag, Pol, Nef, and Env) restricted by the HLA alleles expressed in each subject that contained at least one nonsynonymous mutation for each time point. The rates of escape accumulation were calculated by linear regression modeling, and the statistical uncertainty of these rates was assessed by generating 50 bootstrapped datasets for each patient. This revealed that there was signifi-
cantly greater selection for HLA-associated escape in Env and Nef and less selection for escape in Pol in the VNC group compared with VC children (Fig. 7, A and B). Gag-specific escape did not differ significantly between the two groups. The interquartile range (IQR) of the difference in the rate of escape between VNC and VC children (calculated by bootstrapping viral sequence data from each subject) was estimated in order to assess the statistical robustness of these patterns (Fig. 7 C). These analyses confirm the robustness of the specificity differences of CD8+ T cell selection pressure observed in favor of Nef and Env among VNC children and in favor of Pol among VC children.

Collectively, these data indicate that CD8+ T cells among the slow-progressor children studied exert HLA-associated pressure on the virus from the very first weeks of infection, with unequivocal evidence of escape selection within the first months of life and escape mutants accumulating with age. The patterns of CD8+ T cell activity in the PSPs differ such that in the VNC group, CD8+ T cells preferentially target non-Gag/Pol epitopes and exert greater HLA-associated selection pressure in Nef/Env epitopes and vice versa in the VC children. These distinct patterns of CD8+ T cell activity, together with the observed differences in viral loads in the VC and VNC groups, are consistent with previous studies linking non-Gag/Pol specificity of CD8+ T cell responses with high viremia (Klein et al., 1995; Riviére et al., 1995; Ogg et al., 1998; Brouch et al., 2002, 2003; Edward et al., 2002; Novisky et al., 2003; Masemola et al., 2004; Zuñiga et al., 2006; Geldmacher et al., 2007; Kiepiela et al., 2007; Streeck et al., 2007; Janes et al., 2013), as discussed further below.

**Discussion**

In this study, we sought to investigate the potential for HIV-specific CD8+ T lymphocyte activity to play an effective role in viral eradication of pediatric HIV infection, because CD8+ T cells play a key role in shock-and-kill HIV cure strategies (Deeks, 2012), and because previous studies have indicated that the potential for HIV cure may be higher among infected children than adults (Goulder et al., 2016). Studying ART-naive PSPs followed through childhood, we hypothesized that the breadth, specificity, and function (reflected by the selection of CTL escape mutants) of HIV-specific CD8+ T cell responses would differ according to viremia. We divided these children into low viremia groups (termed VCs and VNCs, respectively) that did not differ significantly by absolute CD4 count or CD4%.

The HIV-specific CD8+ T cell responses in VCs were of lower breadth and magnitude than those of VNCs but focused predominantly on Gag and Pol specificities compared with Nef and other accessory/regulatory proteins (Nef/Acc-Reg) and Env, whereas the reverse applied among VNCs. Unexpectedly, in both groups, escape

**Table 2. HLA-associated mutations in the VC child 133-C that developed within the first months of life**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Child’s age</th>
<th>A*74 Gag</th>
<th>B*81 Pol</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Frequency</td>
<td>Frequency</td>
</tr>
<tr>
<td>133-M</td>
<td>1.1</td>
<td>.E.</td>
<td>100% nd</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>V.E.</td>
<td>100% nd</td>
</tr>
<tr>
<td>133-C</td>
<td>0.3</td>
<td>.E.</td>
<td>100% 65%</td>
</tr>
<tr>
<td></td>
<td>2.7</td>
<td>.V.</td>
<td>75% K</td>
</tr>
<tr>
<td></td>
<td>4.9</td>
<td>.V.</td>
<td>99% K</td>
</tr>
<tr>
<td></td>
<td>7.7</td>
<td>.E.</td>
<td>100% K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HLA association</th>
<th>HLA-A*74</th>
<th>HLA-B*81</th>
</tr>
</thead>
<tbody>
<tr>
<td>133-M</td>
<td>HLA-A*74</td>
<td>HLA-B*81</td>
</tr>
<tr>
<td>Escape variant</td>
<td>K12N</td>
<td>P1895</td>
</tr>
<tr>
<td>q value</td>
<td>6 x 10^{-17}</td>
<td>3 x 10^{-15}</td>
</tr>
<tr>
<td>q value</td>
<td>7 x 10^{-12}</td>
<td>2 x 10^{-10}</td>
</tr>
</tbody>
</table>

Mother 133-M: A*74:01-, B*81:01+. VC child 133-C: A*74:01+, B*81:01+. Footprints associated with the HLA class I allele are indicated below the table; known escape variants and q values indicating significance of an HLA-polymorphism association are shown (Carlson et al., 2014). 133-M indicates mother of the child 133-C. Frequency represents percentage at which a particular haplotype is present at the indicated time point.
mutations were selected from early infancy; however, in VCs, the selection pressure on the virus focused more on Gag/Pol than on Nef/Acc—Reg/Env, and the reverse was the case in VNCs. Finally, preferential recognition of autologous epitope variants in children, but not adults, was shown with respect to Gag-TL9, the immunodominant epitope for approximately one third of HIV-infected individuals in sub-Saharan Africa, also demonstrating the potential for HIV-specific CTL in children to corner the virus.

This escape-cornering strategy has long been proposed as a theoretical immune tactic against HIV (O’Connor et al., 2001; Altfeld and Allen, 2006; Allen and Altfeld, 2008) and has been mathematically modeled (Ferguson et al., 2013). That such a strategy can be effective is demonstrated for the first time here. In the current study, we observe the HLA-B*81:01 TL9-Gag response first drives the selection of an escape variant such as T186M, which substantially reduces viral replicative capacity (Fig. 4); then, by the generation of a variant-specific CTL response, the virus is down an evolutionary cul-de-sac with apparently no further viable escape options. These findings suggest the potential for HIV-specific CD8+ T cells to play an effective role in shock-and-kill HIV-cure strategies among slow-progressor children, because after latency reversal, HIV-specific CD8+ T cells will be required to kill virus-infected cells expressing escape variants in both pediatric and adult infection (Deng et al., 2015).

However, the success of an escape-cornering strategy depends on the ability of the immune system to not only generate variant-specific immune responses as the virus escapes but also drive the selection of escape mutants that significantly diminish viral replicative capacity and limit further escape options. The ability of the immune system to generate sequential immune responses against viral variants and yet fail to corner HIV is especially well described in the context of neutralizing antibody activity against HIV, in which the virus is always one step ahead of contemporary antibody (Wei et al., 2003). Env is highly variable and sequence changes appear to have little impact on viral replicative capacity (Troyer et al., 2009). VNC children whose HIV-specific CD8+ T cell responses are not typically targeting conserved epitopes in Gag and Pol may therefore lack an escape-cornering facility, irrespective of their ability to generate variant-specific CTL responses. To induce such responses may require immunotherapeutic vaccination, ideally using a mosaic insert encoding the common escape variants (Fischer et al., 2007; Barouch et al., 2010), in those conserved epitopes (Hayton et al., 2014; Ondondo et al., 2016). A precedent for the effectiveness of
Table 3. Early escape in the VNC child 517-C develops in non-Gag epitopes

<table>
<thead>
<tr>
<th>Subject</th>
<th>Child's age</th>
<th>C*03 Gag</th>
<th>Frequency</th>
<th>B*42 Nef</th>
<th>Frequency</th>
<th>A*30 Pol</th>
<th>Frequency</th>
<th>A*30 Env</th>
<th>Frequency</th>
<th>B*42 Env</th>
<th>Frequency</th>
<th>B*15:10 Rev</th>
<th>Frequency</th>
<th>B*15:10 Vf</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>517-M</td>
<td>0.25</td>
<td></td>
<td>100%</td>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>20%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>517-C</td>
<td>0.2</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Note: 517-M: A*30.02+ B*42+ B*15:10+ C*03+. Footprints associated with the HLA class I allele are indicated below the table; known escape variants and Q values indicating significance of an HLA-polymorphism association are shown (Carlson et al., 2014). 517-M indicates mother of the child 517-C. Frequency represents percentage at which a particular haplotype is present at the indicated time point.

*HLA-B*15:10-restricted Vif epitope is W19 (WH6GHVS1), residues 79–87; in bold, but HLA-B*15:10-associated polymorphisms are at the residue 85 (V85A) within the epitope and at the residue 78 (D78A) upstream of the epitope.
such an approach in redirecting the specificity of the CTL response toward more conserved regions of the proteome has been described in the Phambili trial of the MRKAd5 Gag/Pol/Nef T cell vaccine, in which induction of higher Gag-specific responses among HLA-B*58:02-positive subjects who became HIV-infected was associated with lower viral loads than in the HLA-B*58:02-positive placebo controls who later became infected (Leitman et al., 2016). In natural HIV infection, HLA-B*58:02 is associated with dominant Env-specific CD8+ T cell responses and high viral loads (Ngumbela et al., 2008). Thus, eradication of the viral reservoir likely requires the ability of HIV-specific CD8+ T cell activity to both recognize autologous variants and operate effectively in killing virus-infected cells, which in part is related to specificity of the response.

The greater ability of infected children compared with adults to generate de novo variant-specific CD8+ T cell responses is described here in relation to the Gag-TL9 epitope that is immunodominant among HIV-individuals expressing HLA-B*42:01 or HLA-B*81:01 and some other HLA class I molecules in the B7 superfamily, such as HLA-B*39:10 (Leslie et al., 2006). These are highly prevalent in the sub-Saharan African populations worst affected by the HIV epidemic, and one or more of these HLA class I molecules are expressed in approximately one third of these individuals. We previously described the same ability of children and not adults to

<table>
<thead>
<tr>
<th>Subject</th>
<th>HLA</th>
<th>Child’s age (yr)</th>
<th>Autologous TPDQLNTML</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-004-M (mother)</td>
<td>Not known</td>
<td>n/a</td>
<td>Not known</td>
</tr>
<tr>
<td>K-004-C (child)</td>
<td>B*81:01+ve</td>
<td>12.8</td>
<td></td>
</tr>
<tr>
<td>K-044-A (mother)</td>
<td>B*42/81-neg</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>K-044-C (child)</td>
<td>B*81:01+ve</td>
<td>4.5</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. Preferential recognition of autologous TL9 variants by HIV-infected children compared with adults. (A) Development of preferential recognition of autologous TL9–Q182G variant by pediatric subject K-004-C between age 8.2 yr, at which time autologous virus encoded a mix of Q182P/Q182G variants, and 12.8 yr, at which time autologous virus encoded only the Q182G variant, by population sequencing of virus. HLA type and autologous sequence of the mother K-004-M are unknown. ELISPOT assays repeated in triplicate. (B) Development of preferential recognition of autologous TL9–Q182G variant by pediatric subject K-044-C between age 4.5 yr, at which time autologous virus encoded a mix of Q182P/Q182T variants, and 8.7 yr, at which time autologous virus encoded only the Q182G variant, by population sequencing of virus. The mother (K-044-M) was HLA-B*42:01/81:01 negative and autologous virus when sequenced at child’s age 4.5 yr encoded wild-type TL9, suggesting that wild-type TL9 was transmitted to the child. ELISPOT assays repeated in triplicate. (C) The magnitude of the IFN-γ ELISPOT response to wild-type TL9 is lower than to the autologous variant among children (n = 11, mean 531 vs. 1,010 spot-forming cells per million PBMCs; P = 0.02, paired t test; P = 0.002, Wilcoxon signed-rank test). Among adults, the IFN-γ ELISPOT response to wild-type TL9 tended to be higher than to the autologous TL9 variant, although this did not reach statistical significance (n = 16; median 1,597 vs. 882 spot-forming cells per million PBMCs; P = 0.25, paired t test; P = 0.40, Wilcoxon signed-rank test). (D) Frequency of subjects showing preferential recognition of autologous variant is significantly higher in pediatric subjects than adult subjects (10/11 = 91%; vs. 5/16 = 31%; P = 0.0047, Fisher’s exact test). SFC, spot-forming cell.
generate autologous variant-specific responses in a study of HLA-B*57/58:01–restricted Gag-TW10 responses, where 73% of children mounted de novo variant-specific responses compared with 9% of adults \((P = 0.0004;\) Feeney et al., 2005). Thus, although the generation of de novo responses to variants is not unique to HIV-infected children (Allen et al., 2005), this phenomenon appears on the basis of this and the current study to be much more common among children. Caveats to this analysis, as mentioned above, include the fact that the viral sequence transmitted by the mother was not known in all cases; however, restricting the analysis to the children in whom either evolution of the variant-specific response could be shown via longitudinal analyses, or in whom maternal virus encoded wild-type Gag-TL9, the preferential recognition of the autologous variant remained significantly more common among children than adults \((P = 0.0045).\) Furthermore, our finding that 5 of 16 (31%) adults had Gag-TL9 variant-specific responses may overestimate the number generating these responses subsequent to a wild-type-specific response, because it is likely that in a proportion of these adults, the virus transmitted encoded a Gag-TL9 variant. Although wild-type virus tends to be preferentially transmitted (Carlson et al., 2016), even when it is present at lower frequency than the escape variant, as illustrated in the current study (for example, Fig. 5 D), as many as 18% of HLA-B*42:01/81:01–negative individuals carry Gag-TL9 variants (Payne et al., 2014), presumably that were selected before transmission.

Why adults might be less able to mount variant-specific responses than children is unknown, but one may speculate that this difference is related to the respective numbers of naive and memory T cells in children and adults. In adults, it may be more likely that a memory CD8+ T cell response would cross-react weakly with a novel CD8+ T cell escape variant; hence, no de novo response would be initiated, giving rise to a situation akin to “original antigenic sin” (Klenerman and Zinkernagel, 1998) among adults. This phenomenon has been proposed to explain disease after dengue reinfection by a distinct subtype of virus from the virus that caused primary infection (Mongkolsapaya et al., 2003).
The observations made here of distinct CD8+ T cell specificities among VC and VNC children are consistent with several studies indicating that Gag- and, to a lesser extent, Pol-specific CD8+ T cell responses are more effective at controlling viremia than non-Gag/Pol specificities (Klein et al., 1995; Riviére et al., 1995; Ogg et al., 1998; Barouch et al., 2002, 2003; Edwards et al., 2002; Novitsky et al., 2003; Mason et al., 2004; Zúñiga et al., 2006; Geldmacher et al., 2007; Kiepiela et al., 2007; Sacha et al., 2007a,b; Streeck et al., 2007; Kawashima et al., 2009, 2010; Payne et al., 2010; Goulder and Walker, 2012; Janes et al., 2013). The mechanisms for these protein-specific differences remain unclear but may be related to the ability of Gag and Pol epitopes in particular to be presented on the surface of infected cells before de novo HIV protein synthesis and concomitant HLA down-regulation (Sacha et al., 2007a,b) and substantially earlier than Nef or Env specificities. However, it is evident that factors other than protein specificity, including functional avidity and polyfunctional- ity (Almeida et al., 2007, 2009), also play an important role in the ability of a particular CD8+ T cell response to control HIV replication. Factors contributing to the differences in specificity of the HIV-specific CD8+ T cell responses in the VNC and VC groups of children are likely to include those shown to be important in adult infection, namely the HLA class I molecules expressed (Kiepiela et al., 2004) and the viral sequence transmitted (Goepfert et al., 2008; Carlson et al., 2016). These factors have previously been shown to also have a strong influence on specificity of the CD8+ T cell response in infected infants (Goulder et al., 2001b; Thobakgale et al., 2007). The impact of transmission of a virus preadapted to HLA in the child will be especially marked, because the mother shares at least half the class I molecules expressed in the child, and in such cases, the transmitted virus may be highly preadapted (Goulder et al., 2001b; Thobakgale et al., 2007; Goepfert et al., 2008; Carlson et al., 2016). Notwithstanding these differences in the protein specificity of the CD8+ T cell response and in the viral loads of the two subgroups of PSPs studied here, these were not reflected in CD4 count differences (Fig. 1), underlining the lack of a significant correlation typically observed between viral load and absolute CD4 count in pediatric infection (Sewanyana et al., 2007; Muenchhoff et al., 2016).

The observation of high rates of CD8+ T cell escape in Nef and Env in the VNC pediatric subjects is consistent with the dominant Env- and Nef-specific CD8+ T cell responses detected. Previous studies in adults have shown that rate of escape is related to the magnitude of the response, its relative immunodominance compared with the other responses generated, and epitope entropy (Ferrari et al., 2011; Liu et al., 2013). The rate and location of escape mutations is to a surprisingly large degree predictable, independent of the CD8+ T cell response actually generated (Barton et al., 2016), based on knowledge of epitopes, the transmitted viral sequence, and entropy. Although the mean entropies of Gag and Pol proteins are substantially less than those of Env and Nef (Yusim et al., 2002), and much of the analysis here has focused on protein-specific CD8+ T cell responses and immune escape, it is important to note that there are also relatively conserved regions within Nef and Env and that the impact of escape mutations on viral replicative capacity may differ greatly between epitopes within the same protein. Indeed, the impact of an escape mutation may be vastly different, depending on which particular mutation is selected within a single epitope. This is well illustrated by the Gag-TL9 epitope studied here and previously (Fig. 4; Wright et al., 2010, 2012; Tsai et al., 2016), in which the T186S and T186M variants such as Q182S have relatively little impact.

The observation here of widespread early HIV-specific CD8+ T cell escape in pediatric infection was unexpected given previous studies that have shown a striking lack of efficacy of anti-HIV CD8+ T cell responses in infected children, even involving the same specificities that are believed to underlie HLA class I-mediated control of viremia in adult infection. For example, the HLA-B*57/58:01-restricted Gag TW10 epitope (Gag residues 240–248) is one of the earliest HIV-specific CD8+ T cell responses generated in acute infection (Altfeld et al., 2001, 2006; Leslie et al., 2004; Brumme et al., 2008) and characteristically drives the selection of escape mutants such as T242N that reduce viral replicative capacity (Martinez-Picado et al., 2006) or, more rarely, other variants such as A248D that substantially cripple the virus (Miura et al., 2009). However, even though a high-frequency HLA-B*57/58:01-restricted TW10 response can be evident at birth among in utero-infected infants, progression is observed without any selection of escape mutants, suggesting lack of antiviral efficacy in this age group (Adland et al., 2015). In contrast, among HLA-B*57/58:01 adults with disease progression, TW10 escape is almost universal (Martinez-Picado et al., 2006). Furthermore, the absence of HLA-B*57/58:01-restricted TW10 response can be mediated protection against disease progression in pediatric infection, contrasting with the strong influence of these alleles in adult infection, is additional indirect evidence that HIV-specific CD8+ T cell responses play an insignificant role in pediatric slow progression (Adland et al., 2015). Finally, more recent data directly measuring magnitude and breadth of virus-specific CD8+ T cell activity further indicate that these responses do not contribute significantly to progression rates in pediatric infection (Muenchhoff et al., 2016). Indeed, a progressively more active CD8+ T cell response that is observed through childhood (Luzuriaga et al., 1995; Scott et al., 2001; Thobakgale et al., 2007) can be associated with higher levels of immune activation and accelerated CD4 decline in these children (Muenchhoff et al., 2016).

The current studies demonstrate the benefit of next-generation deep sequencing (NGS) compared with previous approaches involving population sequencing and/or sequencing of limited numbers of viral clones to address the question of early escape in pediatric infection have (Leslie et al., 2004; Pillay et al., 2005; Sanchez-Merino et al., 2005).
Here, we have been able to demonstrate via NGS the absence of rare variants in the transmitting mother. This is important, because even if the mother does not carry the HLA allele driving selection pressure in the child, it is not uncommon for the virus that is transmitted to the mother to have come from the father of the child and hence to include variants within epitopes that are restricted by HLA only shared by the father and the child (Pillay et al., 2005). The combination of NGS showing no variant in the transmitting mother, and 100% selection of a variant within an epitope restricted by an HLA allele only expressed in the child, as illustrated by the HLA-A*74–associated variant K12N (Table 2), is compelling evidence of early selection pressure on the virus driven by CD8⁺ T cell responses among these infants.

The significance of these findings for slow-progressor children who have gone onto ART as per the guidelines recommending universal ART for all HIV-infected children, irrespective of health of CD4 count, has been discussed above. HIV-specific CTL responses in VC children have the potential to make an effective contribution to eradication of viral reservoir after latency reversal, both because of the capacity to recognize autologous variants and because the CTL specificities are generally Gag/Pol specific and effective at killing virus-infected cells. HIV-specific CTL responses in VNC children may require redirection toward these more conserved Gag/Pol specificities via immunotherapeutic immunization in order to be effective. For the HIV-infected children not studied here in whom ART was initiated early, either prophylactically to prevent immune decline or therapeutically to allow immune reconstitution, it is possible that the viral reservoir may not be comprised so extensively with escape variants because of lack of selection pressure as described above. However, it is likely that immunotherapeutic immunization via a T cell vaccine inducing broad, conserved Gag/Pol specificities in these children also would increase the likelihood of HIV-specific CTL making an effective impact on the viral reservoir.

In considering how the situation might differ between HIV-infected adults and children, several factors may contribute. In HIV-infected adults, widespread CD8⁺ T cell escape typically occurs early in infection (Gouler and Walker, 1999; Brumme et al., 2008; Goonetilleke et al., 2009; Boutwell et al., 2013), and most of the latent reservoir virus expresses escape mutants (Deng et al., 2015). Thus, the relative inability of the adult immune system to mount autologous variant-specific CD8⁺ T cell responses may represent a major obstacle to effective clearance of the adult viral reservoir. Immune reconstitution is both more rapid and more complete in children than in adults on ART (Franco et al., 2000; Gibb et al., 2000; Douek et al., 2003; Feneey et al., 2003; Walker et al., 2004; Sabin et al., 2008), and therefore, the immunogenicity of a T cell vaccine may be substantially greater in ART-treated children (Swadling et al., 2016). Immunotherapeutic vaccination using a mosaic insert encoding the common escape variants (Fischer et al., 2007; Barouch et al., 2010) might be expected to induce responses to the autologous variants more effectively in children than in adults. In all these respects, therefore, the HIV-specific CD8⁺ T cell responses in infected children may be more likely than those in adults to play an effective part in eradication of the viral reservoir. If CD8⁺ T cell responses against “low fitness” variants can be induced either naturally or via vaccination in children, the prospects for HIV “remission” after ART (Persaud et al., 2013; Sáez-Cirión et al., 2013; Frange et al., 2016) may be greater in pediatric than adult infection.

MATERIALS AND METHODS

Study subjects

Eleven children studied here were from the previously described Pediatric Early HAART and Strategic Treatment Interruption Study (Mphatswe et al., 2007; Thobakgale et al., 2007, 2009; Prendergast et al., 2008). This feasibility study enrolled 75 HIV-infected infants born to HIV-positive mothers at St. Mary’s and Prince Mshiyeni Hospitals in Kwa-Zulu-Natal, South Africa, in 2003–2005. Antenatal mothers were recruited in the third trimester of pregnancy; mothers and infants received a single dose of nevirapine at delivery (the only available regimen for prevention of mother-to-child transmission at the time; infant ART was not available at government hospitals). Infants were randomized to one of three study arms: deferred treatment, immediate uninterrupted treatment for 12 mo, or immediate 18-mo treatment with structured interruptions (Mphatswe et al., 2007; Prendergast et al., 2008). Diagnosis of HIV infection in infants and HLA class I typing were performed as previously described (Kiepiela et al., 2004; Mphatswe et al., 2007; Thobakgale et al., 2007; Prendergast et al., 2008). CD4⁺ T cell counts were determined by flow cytometry using standard clinical protocols. HIV plasma viral load measurements were done using the Roche Amplicor Monitor assay according to the manufacturer’s instructions. Eleven children from this cohort were termed PSPs, defined here as children infected via mother-to-child transmission who before universal ART recommendations for pediatric HIV infection had not met ART initiation criteria by the age of 5 yr (CD4 >350 cells/mm³ or WHO stage III or IV; Table 1).

The study was approved by the Biomedical Research Ethics Committee of the University of KwaZulu-Natal and the Institutional Review Boards of the Massachusetts General Hospital and the University of Oxford. The mothers gave written informed consent for participation of their children.

Additionally, for the analysis of CD8⁺ T cell responses to the wild-type versus autologous variant-specific epitopes (Fig. 6), we used PBMCs from treatment-naive HIV C clade–infected subjects from the following previously described cohorts: pediatric Kimberley cohort from Kimberley, South Africa (Adland et al., 2015); adult Gateway cohort from Durban, South Africa (Payne et al., 2014); and adult Thames Valley Cohort from Northampton General Hospital,
Northampton, UK, and Wycombe Hospital, High Wycombe, UK (Payne et al., 2010). All were C clade–infected adults with chronic HIV infection. Among the Kimberley pediatric subjects, the mean age at enrolment was 6.5 yr, mean CD4 count was 876 cells/mm³, mean CD4% was 31.4%, and median viral load was 33,000 copies/ml. Among the adults, the mean CD4 count was 531 cells/mm³, (IQR, 429–576 cells/mm³), and the median viral load was 8,700 HIV copies/ml plasma (IQR, 5,090–27,814). All participants provided written informed consent. Ethics approval was given by the University of KwaZulu-Natal Biomedical Research Board and the University of Oxford Ethics Committee.

IFN-γ ELISPOT assays
Ex vivo HIV-specific CD8⁺ T cell responses were screened in IFN-γ ELISPOT assays using a panel of 410 overlapping 18mer peptides, spanning the entire HIV proteome, based on 2001 C clade consensus sequence, as described previously (Addo et al., 2003; Kiepiela et al., 2004; Thobakgale et al., 2007). Spots were counted using an automated ELISPOT reader (AID ELISPOT v4.0; Autoimmun Diagnostika) and manually checked. Background (>3 SDs above the mean of the 4 wells containing PBMCs without a peptide) was subtracted from values of all wells, and the final result was expressed as number of spot-forming cells per 10⁶ PBMC. Positive responses were considered >100 spot-forming cells/10⁶ PBMC after background subtraction. In IFN-γ ELISPOT assays to determine recognition of wild-type TL9 and autologous Gag-TL9 variants, where adequate cell numbers were available, assays were performed in triplicate at peptide concentrations of 10⁻⁶ M to 10⁻⁴ M (Fig. 6, A and B). In subjects where cell numbers were limiting, peptide concentrations of 10⁻⁶ M to 10⁻⁴ M were used. Data shown in Fig. 6 (C and D) were determined by calculating the mean of the responses to TL9 wild-type and autologous TL9 variant peptides at concentrations of 10⁻⁶ M to 10⁻⁴ M for each subject.

Viral RNA extraction
One-milliliter plasma aliquots were thawed at 37°C in a water bath and centrifuged for 1.5 h at 15,000 rpm; 860 µl supernatant was removed, and viral RNA was isolated from the remaining 140 µl using QIAmp Viral RNA Mini kit (QIAGEN) following the manufacturer’s instructions. Samples with a viral load <2,000 copies/ml were concentrated by putting two or three aliquots of plasma (if available) through the same QIAmp column.

Population sequencing of gag
Genomic DNA was extracted from whole blood via the QIAamp DNA Blood Mini kit (QIAGEN). HIV gag sequences were amplified by nested PCR to obtain population sequences using the primers 5’-CTCTAGCATGTCCTCGCAG CCGAA-3’ and 5’-TCTCTTTCCACATTTCCAACAG CC-3’ for the first round and 5’-CAATTCTGGCTATG TGCCC-3’ and 5’-ACTCAGGCTTGATGAAAGTGC-3’ for the second round. Viral RNA was isolated from plasma by use of a QIAamp Viral RNA Mini kit (QIAGEN). The Gag-Protease region was amplified by RT-PCR from plasma HIV-1 RNA using Superscript III One-Step Reverse transcription kit (Invitrogen) and the following Gag-protease-specific primers: 5’-CAGTGTAGGCCCTAATAAAAGCT TGCC-3’ (HXB2 nucleotides 512–539) and 5’-TTTAAC CCTGTGCTTGTGTTATTCTCT-3’ (nucleotides 2,851–2,825). Second-round PCR was performed using forward primer (5’-GACTCGGCTTTGCTGAAGCGCAGCCACGGG AAGAGCCAGGGCGACCTGGTGAAGTGACCGC AATTAGCTAGGCGAGGTAGAAGAGAGAGA TGCG-3’, 695–794) and reverse primer (5’-GGCCCAATT TTGAAATTTTTCCCTTTTTTTGTA CAAAATCTACAAATGCTTATTTTCTGTC AATGCGATTGTTAACCTTG-3’, 2,646–3,547). All PCR products were purified using a QIAquick PCR purification kit (QIAGEN) according to manufacturer’s instructions. Sequencing was undertaken using the Big Dye Ready Reaction Terminator Mix (V3; Applied Biosystems) analyzed using Sequencer v4.8 (Gene Codes Corporation) and manually aligned using Se_Al software.

Amplification of full-length viral RNA for ultra-deep sequencing
Amplification of the full HIV genome in four overlapping fragments was performed using SuperScript III One-Step RT-PCR System with Platinum Taq High Fidelity (Invitrogen), as previously described (Gall et al., 2012). To confirm amplification, PCR products were run on 1% agarose gel with 240 V, 300 mA for 1.5 h, and bands were visualized under UV light. Amplicons were then combined into one pool (5 µl of amplicon 1 and 10 µl each of amplicons 2, 3, and 4) and ultra-deep sequenced at the Wellcome Trust Sanger Institute (Cambridge, UK).

Ultra-deep sequencing and de novo assembly of viral genomes
Libraries were prepared from 50–1,000 ng DNA as described previously (Quail et al., 2008, 2011) using one of 96 multiplex adaptors for each pool of amplicons. Paired-end sequencing with a read length of 300 bp was performed using the Illumina MiSeq instrument as described previously (Gall et al., 2014). Consensus sequences were generated by de novo assembly (i.e., without a reference sequence) using Iterative Virus Assembler (Hunt et al., 2015).

Epitope haplotype calling from ultra-deep sequencing data
Gaps in the de novo assembled genomes were filled using maternal or a reference sequence. The reads were then mapped to the assemblies using MOSAIK (Lee et al., 2014) with default parameters. A subtype C reference genome was annotated with the location of the epitopes of interest. The de novo assemblies were aligned using MUSCLE (Edgar, 2004).
to the annotated reference to determine the location of the epitopes in each assembly. V-Phaser 2 with default parameters was used to call variants (Yang et al., 2013) and V-Profiler to call haplotypes (Henn et al., 2012) for the HLA-associated epitopes. Only haplotypes with greater than 1% frequency were included in the analysis.

Phylogenetic analysis of HIV genomes
HIV genomes from 11 PSP mother–child pairs were ultra-deep sequenced. Maternal time points included closest to delivery, if available, and 5–7 yr after delivery, if available; children’s time points included earliest available, ~2.5, ~5, and ~7 yr of age or only the earliest and 5 yr of age time points (43 full genomes and seven partial genomes were generated). To confirm the close relationship between viral sequences from mother–child pairs, a maximum-likelihood phylogenetic tree was constructed using the General Time Reversible models of nucleotide substitution, from the 43 full-genome sequences and eight clade-specific reference sequences with 1,000 bootstrap replicates in MEGA version 6.06-mac software (Tamura et al., 2013). The tree was visualized in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) and Adobe Illustrator CS4 v14.0.0. HIV full-genome sequences and eight clade-specific reference sequences were then translated in Geneious 6.1.8 and aligned in Mega6.06. A panel of well-characterized epitopes (Table S4) across full HIV proteome restricted by the six HLA-A/B/C alleles expressed among the 11 PSPs was compiled. To examine escape in HLA-relevant CD8+ T cell epitopes, for each child, regions of the virus encoding the CD8+ T cell epitopes restricted by any of the six HLA alleles expressed by that child were analyzed. Only polymorphisms that developed independently in the child (i.e., had not been transmitted by the mother) were included and confirmed using epitope minor variant data.

The extent of HLA-driven escape was quantified as the proportion of HLA-associated sites that were adapted to that allele (i.e., polymorphisms that are enriched in the presence of an HLA allele or polymorphisms that differ from the polymorphisms that are less frequently observed in the presence of that HLA allele) based on previously defined HLA-associated escape sites (Carlson et al., 2012, 2014). The number of HLA-associated adaptations observed across the HIV proteome in each child was divided by the total number of HLA-associated epitopes relevant to that child’s HLA repertoire (Payne et al., 2014).

Calculating the contribution to escape in VC and VNC children targeting Gag-TL9
The Gene Cutter tool (http://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html) was used to extract sequences of the nine HIV proteins from the de novo assembled consensus sequences from mother–child pairs; protein-encoding sequences were then translated in Geneious 6.1.8 and aligned in Mega6.06. A panel of well-characterized epitopes (Table S4) across full HIV proteome restricted by the six HLA-A/B/C alleles expressed among the 11 PSPs was compiled. To examine escape in HLA-relevant CD8+ T cell epitopes, for each child, regions of the virus encoding the CD8+ T cell epitopes restricted by any of the six HLA alleles expressed by that child were analyzed. Only polymorphisms that developed independently in the child (i.e., had not been transmitted by the mother) were included and confirmed using epitope minor variant data.

The extent of HLA-driven escape was quantified as the proportion of HLA-associated sites that were adapted to that allele (i.e., polymorphisms that are enriched in the presence of an HLA allele or polymorphisms that differ from the polymorphisms that are less frequently observed in the presence of that HLA allele) based on previously defined HLA-associated escape sites (Carlson et al., 2012, 2014). The number of HLA-associated adaptations observed across the HIV proteome in each child was divided by the total number of HLA-associated epitopes relevant to that child’s HLA repertoire (Payne et al., 2014).

Calculating the rate of escape in HLA-restricted CD8+ T cell epitopes
We have defined escape as a nonsynonymous mutation that has arisen in the HLA-restricted CD8+ T cell epitope sequence since the founder virus, which we infer from the consensus sequence of the first time point. A panel of well-characterized epitopes (Table S4) across full HIV proteome restricted by six HLA-A/B/C alleles expressed among the 11 PSPs was compiled. To examine the accumulation of escape mutations in different gene regions, the proportion of sequences that accumulated at least one escape mutation was calculated for each time point for known restricted CD8+ T cell epitopes per patient. Furthermore, we ensured that there were at least 100 sequences with 95% coverage of the epitope region to estimate the proportion of “escaped” sequences. The mean accumulation of escaped variants per gene was obtained by fitting a linear regression model to the proportion of escaped variants over time. We also explored a logistic regression model, although the sum of squared errors and Akaika information criterion indicated that the linear regression model provided a significantly better fit to the data. The statistical uncertainty of these rates was assessed by generating 50 bootstrapped datasets for each patient. The difference in the rates of escape was estimated by calculating the difference in the slope of the linear regression model (per bootstrap replicate) between VNC and VC children for each gene. These analyses were performed using custom-made python scripts (code available at https://github.com/jnarag/DeepGenomeAnalysis).

Site-directed mutagenesis of T186M
T186M mutation of HIV-1 Gag sequence was introduced into SK-254(M), a modified version of a patient-derived subtype C HIV-1 Gag-protease sequence (SK-254, GenBank accession no. HM593258) with p24 Gag identical to consensus C inserted in pNL4-3. The mutation was engineered by using QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies) along with custom-designed mutagenesis forward and reversed primers: 5’-CACCCCA AGATTAAAAACATGATGTTAAATACAGTGGGGGG-3’ and 5’-CCCCCCCACGATTAAAATACAGTGGGGGG-3’. The mutation was confirmed by sequencing after the mutagenesis.

Virus propagation
Plasmid preparation was performed according to manufacturer’s instructions (HiSpeed plasmid Maxi kit; QIAGEN). To generate mutant viruses, the T186M Gag-Pro amplified and purified PCR products and BstE II linearized pNL4-3Δgag-protease (New England Biolabs) were transfected into the GFP-reporter GXR cell line via electroporation in a Bio-Rad GenePulsar II using 0.4-cm cuvettes at 300 V, 500 μF and infinite resistance as previously described (Miura et al., 2009). Virus propagation was monitored by flow cytometry to detect infected GFP-positive cells. Virus culture
supernatants were harvested when 30% of cells were GFP-positive. Viruses were aliquoted and stored at −80°C until use.

Statistical analysis
Statistical analyses were performed in Prism (v5.0c; GraphPad) or R (v3.3.1; The R Foundation for Statistical Computing). Linear mixed effect regression analyses were performed using the lme4 package in R, with the lmer (for continuous outcomes) or glm (for binomial outcomes such as PSP group) functions, in each case we allowed for a random effect attributable to the age in days at which each donor was sampled, and solved for the (fixed) effect estimate caused by PSP group. When estimating the change in immune response over time (as in Fig. 2), we modeled age as a fixed effect and allowed for a random effect caused by donor. We calculated likelihood ratio p-values using the ANOVA function in R to compare two nested models (i.e., a model with the predictor variable of interest versus a null model omitting only that variable) that were fit under a maximum-likelihood scenario. The p-values reported are not adjusted for multiple comparisons. Data plots were made using the ggplot2 package, and the default span (0.75) was used for the LOWESS smoothed lines shown in the figures. The code used for analyses is available from the authors.

For comparisons between two groups at a single time point, or of mean measures, data were analyzed in Prism. Data were first confirmed to have parametric or nonparametric distribution (D’Agostino and Pearson omnibus normality test). For two-group analyses $t$ test (parametric) or Mann–Whitney $U$ test (nonparametric) was performed; for >three-group analyses ANOVA (parametric) or Kruskal–Wallis (nonparametric) followed by post-hoc test was performed. Strength of association between two variables was analyzed by Pearson (nonparametric) or Spearman (parametric) correlation. P-values < 0.05 were considered significant.

Online supplemental material
Fig. S1 shows phylogenetic tree from 43 full-genome sequences of the 11 PSP and their mothers. Table S1 shows HLA-associated escape mutations that developed in the VNC child 517-C. Table S2 shows HLA-associated escape mutations that developed in the VNC child 021-C. Table S3 shows 11 HLA-B*42:01-positive and/or HLA-B*81:01-positive children from Southern Africa in whom autologous virus encoded a Gag-TL9 variant. Table S4 lists epitopes used in the analysis of pediatric deep sequence data. Tables S1–S4 are provided as Excel files.

ACKNOWLEDGMENTS
This work was supported by the Welcome Trust (WT 104748MA to P.J.R. Goulder, 098051 to P. Kellam and A. Gall, and 093768/Z/15/X to A.J. Prendergast); the Clarence Mochida Memorial Foundation for Medical and Pharmaceutical Research (M. Mori).

The authors declare no competing financial interests.

Author contributions: EM. Leitman designed the study, performed experiments, analyzed data, and wrote the manuscript. C.F. Thobakgale and A. Adland performed experiments and analyzed data. M.A. Ansari analyzed data. J. Raghwani and O.G. Pybus analyzed data and contributed to manuscript writing. A.J. Prendergast, G. Tudor-Williams, P. Kiepiela, J. Hemelaar, P. Jooste, T. Ndugu’u, L. Riddell, and G. Luzi managed cohort recruitment and reviewed the manuscript. J. Brenner, M.-H. Tsai, and M. Mori contributed to experiment performance and reviewed the manuscript. B.D. Walker contributed to study design and reviewed the manuscript. P. Kellam and A. Gall performed some of the experiments, contributed to data analyses, and reviewed the manuscript. V. Narain analyzed data and contributed to manuscript revision. P.C. Matthews designed and supervised the study, contributed to data analyses, and wrote and reviewed the manuscript. P.J.R. Goulder conceived, designed, and supervised the study, contributed to data analyses, found funding for the study, and wrote the manuscript.

Submitted: 15 December 2016
Revised: 22 June 2017
Accepted: 30 August 2017

REFERENCES


