

**Identification of HIV Mutations in Donors Experiencing Virologic Failure on Raltegravir-  
Containing Regimens**

Cailyn Barthlow

BIOL 499 Departmental Honors

Hood College

April 25, 2022

## **Identification of HIV Mutations in Donors Experiencing Virologic Failure on Raltegravir-Containing Regimens**

Cailyn Barthlow<sup>1</sup>, Ann Wiegand<sup>1</sup>, Yuta Hikichi<sup>1</sup>, Wei Shao<sup>2</sup>, Rachel Sklutuis<sup>1</sup>, Valerie Boltz<sup>1</sup>, Brandon Keele<sup>3</sup>, John Mellors<sup>4</sup>, Eric Freed<sup>1</sup>, Mary F. Kearney<sup>1</sup>, Jennifer L. Groebner<sup>1</sup>

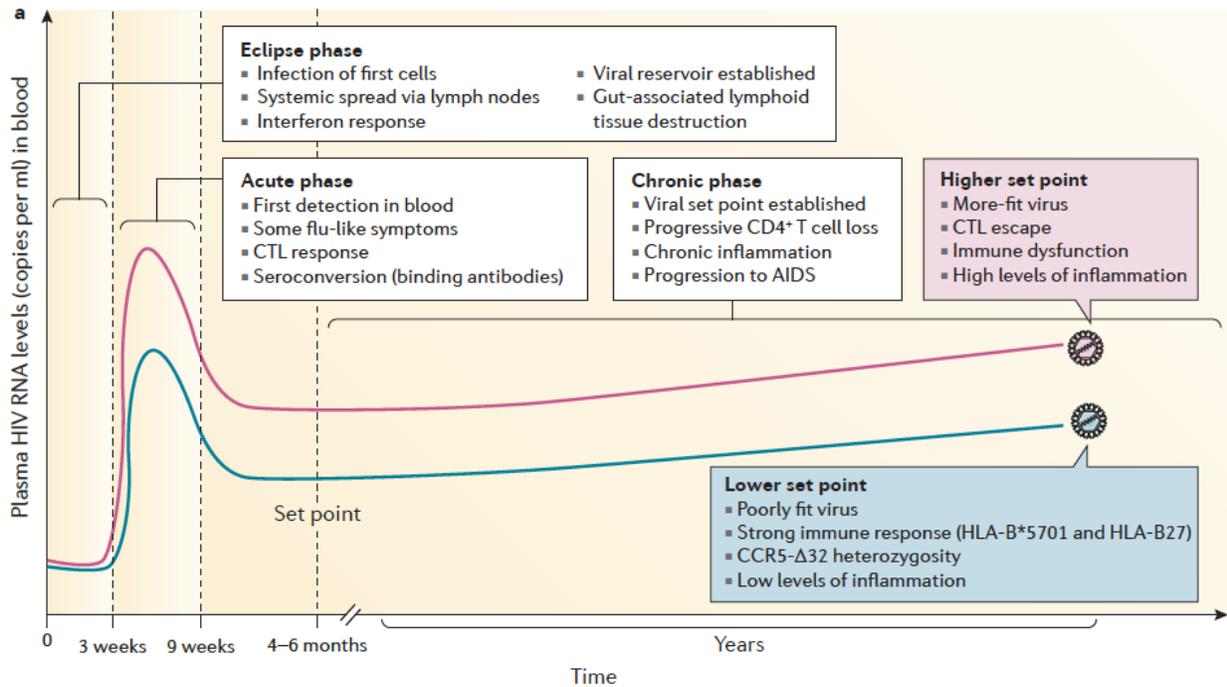
<sup>1</sup>HIV Dynamics and Replication Program, National Cancer Institute, Frederick, Maryland;

<sup>2</sup>Advanced Biomedical Computing Science, Frederick National Laboratory for Cancer Research (FNLCR) sponsored by the National Cancer Institute, Frederick, Maryland.; <sup>3</sup>AIDS and Cancer Virus Program, Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA.;

<sup>4</sup>Department of Medicine, University of Pittsburgh, Pittsburgh, PA 15213

### **Introduction**

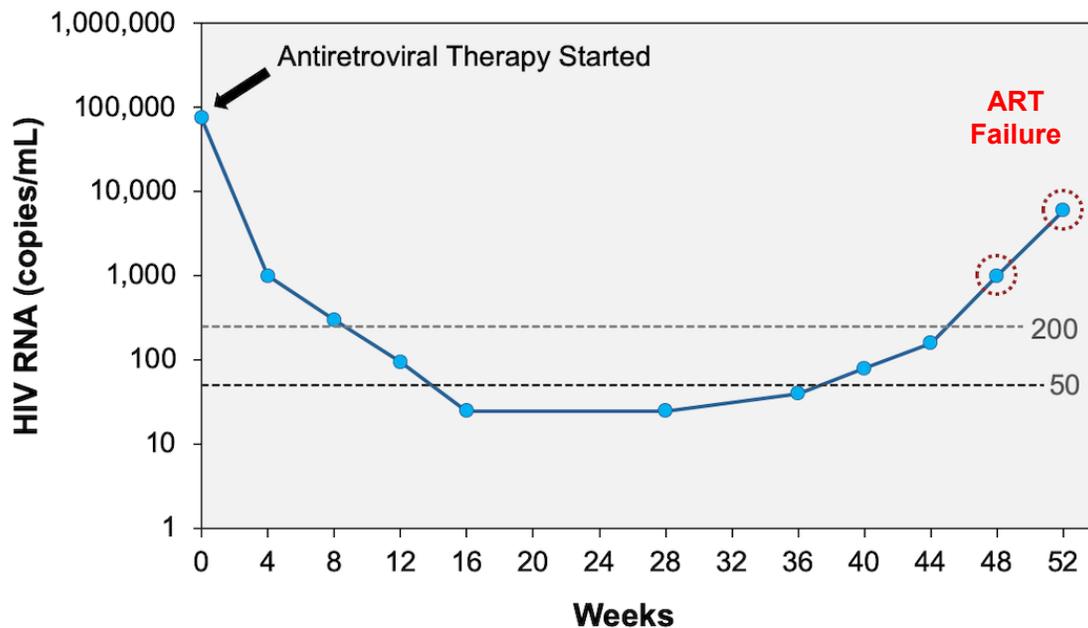
Since the first report of acquired immunodeficiency syndrome (AIDS) in 1981 and the discovery of human immunodeficiency virus (HIV) as the cause in 1983, the HIV/AIDS pandemic has resulted in 36 million deaths worldwide (HIV.gov, 2021). In 2020, about 38 million people were living with HIV, 1.5 million people were newly infected, and 680,00 people died from AIDS-related illnesses (WHO, 2022). While there is still no cure, combination antiretroviral therapy (ART) is an effective treatment and as of June 2021, 28.2 million people were accessing ART, but many individuals still lacked access particularly in resource-limited countries (WHO, 2022).



**Figure 1.** HIV infection. Eclipse phase, acute phase and chronic phase are illustrated with their corresponding HIV RNA levels over time (Deeks et al., 2015).

HIV is a retrovirus that primarily infects CD4<sup>+</sup> T cells and integrates its DNA in the host genome. In untreated individuals, the infection progresses in three stages (Figure 1). The eclipse phase during the first 3 weeks of infection, is characterized by HIV infecting cells in mucosal tissue, spreading through the lymphoid system and establishing the viral reservoir as early as 3 days after infection (Deeks et al., 2015). The viral reservoir is defined as intact, replication-competent virus that accumulates and persists in cell types or anatomical sites in the body. From weeks 3 to 9 post-infection, exponential replication and spread of the virus results in the virus becoming detectable in the blood during the acute phase. During this phase, HIV RNA levels in the blood peak around day 30 and gradually, the adaptive immune response achieves some control of the virus (Deeks et al., 2015). However, the virus is not cleared by the immune response and plasma virus levels are maintained at a setpoint during the chronic phase which begins between ~4-6 months post-infection. In individuals not on ART, HIV plasma RNA levels

persist due to continuing rounds of viral replication and infection of new cells, resulting in the progressive loss of CD4<sup>+</sup> T cells. Over time, progression of disease leads to AIDS with higher viral loads and higher levels of inflammation and immune dysfunction, which increases the risk of individuals to develop opportunistic infections and cancers like Kaposi's sarcoma (Deeks et al., 2015).

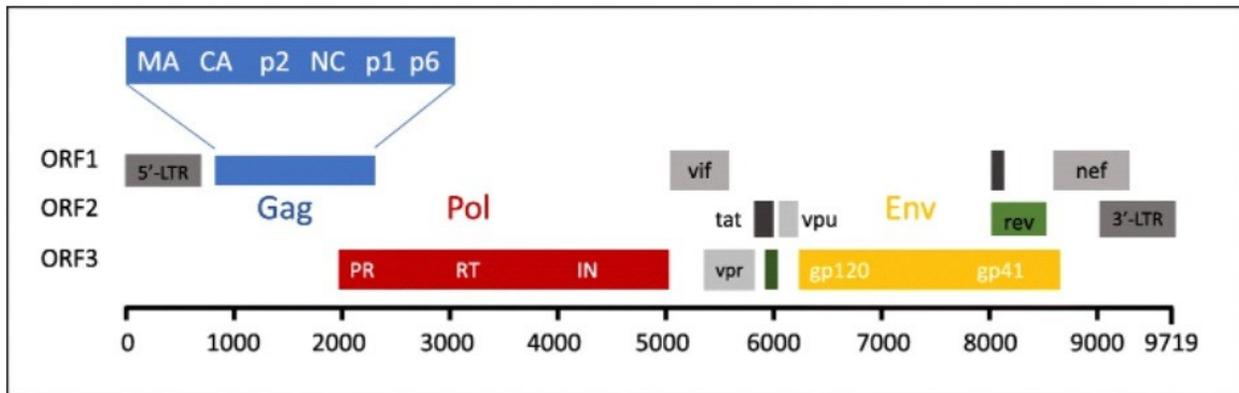


**Figure 2.** HIV plasma HIV-1 RNA levels following ART initiation and during ART failure (Spach, 2022).

When individuals initiate ART (Figure 2), on-going viral replication and infection of new cells is halted (Kearney et al., 2014; Lee et al., 2019; McManus et al., 2019; Van Zyl et al., 2017) and plasma viral RNA levels decline below clinically detectable levels (<50 copies/ml) to a viral setpoint of on average 1 HIV RNA copy/mL (Palmer et al., 2008). If ART is interrupted, rebound in viremia to pre-ART levels occurs (Finzi et al., 1997, 1999; Wong et al., 1997). Occasionally, drug resistance mutations can emerge or be selected within the HIV population,

leading to increases in plasma viremia. This rebound is known as ART failure and occurs when plasma viral RNA levels exceed 200 copies/mL.

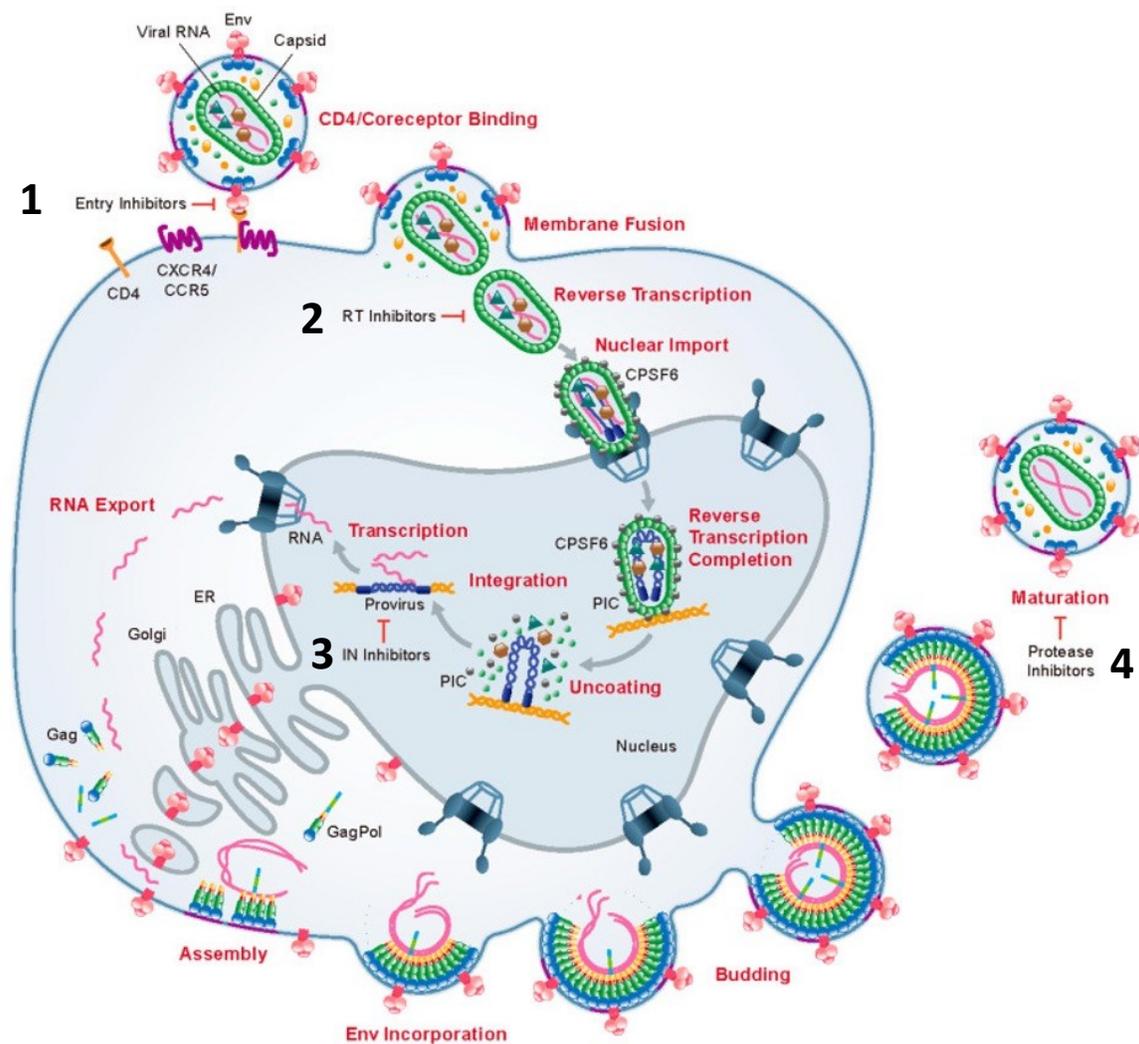
The HIV genome is approximately 9,200-9,600 nucleotides and consists of 9 genes encoding 15 viral proteins (Figure 3). The HIV provirus (the DNA form of the virus) is flanked by LTRs (long terminal repeats) which contain regulatory regions important for transcription and polyadenylation. The structural proteins and viral enzymes are the products of *gag*, *pol*, and *env* genes. *Gag* encodes the capsid proteins: Matrix (MA), Capsid (CA), Nucleocapsid (NC) and p6. MA plays a crucial role of signaling *gag* proteins to the plasma membrane and in the integration of *env* into the virion during assembly (Göttlinger, 2002). CA forms a conical shell around the viral RNA in the virion and plays a role in particle assembly (Göttlinger, 2002). NC is responsible for packaging the viral RNAs during assembly (Göttlinger, 2002). Lastly, the p6 protein plays a role in facilitating viral release from the cell by recruiting cellular factors (Göttlinger, 2002). The *pol* gene encodes the viral enzymes protease (PR) that cleaves of the *gag-pol* polyprotein, reverse transcriptase (RT) which reverse transcribes the RNA into complementary DNA and integrase (IN) that is responsible for integrating the viral DNA into the host genome (Foley et al., 2018). *Env* encodes the viral glycoproteins, gp120 and gp41 which are important in binding to the receptor and coreceptor on the target cell, and for fusion of the viral and cellular membranes (Foley et al., 2018). HIV also contains the regulatory proteins, Tat and Rev, that are crucial for transcriptional and posttranscriptional processes of viral gene expression. Additional HIV proteins are the accessory proteins which includes, Vpr, Vpu, Vpx and Nef (Los Alamos National Library, 2017). Nef specifically causes the downregulation of the CD4 receptor and enhances viral infectivity and disease progression (Los Alamos National Library, 2017).



**Figure 3.** HIV-1 Proviral DNA genome consists of 9 viral genes and 15 proteins (Cervera et al., 2019).

HIV-1 infects cells (Figure 4) facilitated by gp120 of the viral envelope binding to the CD4 receptor followed by a conformational change allowing gp41 to bind to the coreceptor, CCR5 (Flint et al., 2020). This interaction induces further conformational changes in the envelope glycoprotein where cleavage exposes the fusion domain in gp41 resulting in fusion of the viral and target cell membranes (Flint et al., 2020). The viral capsid is released into the cell and the positive single stranded RNA genome is converted to DNA by reverse transcriptase as the capsid is trafficked to the nucleus. In the nucleus, the capsid uncoats and reverse transcription is completed. The viral DNA is then integrated into transcriptionally active sites of the host cell genome coordinated by the viral enzyme, integrase (Kleinpeter & Freed, 2020). The integrated proviral DNA is transcribed to produce viral mRNAs and full-length genomic RNA (Kleinpeter & Freed, 2020). Full length mRNA makes Gag and Gag-Pol polyproteins at about a 20:1 ratio and these are sent to the plasma membrane where they begin to assemble the hexameric lattice structure of the virion (Kleinpeter & Freed, 2020). The envelope glycoprotein precursor, gp160, is trafficked to the ER-Golgi complex where it is cleaved by host furin proteases resulting in the mature envelope glycoproteins, gp120 and gp41 (Flint et al., 2020). These glycoproteins are then trafficked to the plasma membrane where they are embedded to form the envelope of the virus

(Flint et al., 2020). The viral enzymes and viral genome are packaged into the virion and the immature virus particle buds off the host cell (Flint et al., 2020). After release, protease cleaves Gag polyproteins into their mature proteins allowing the conical capsid to form creating the mature virion (Kleinpeter & Freed, 2020).

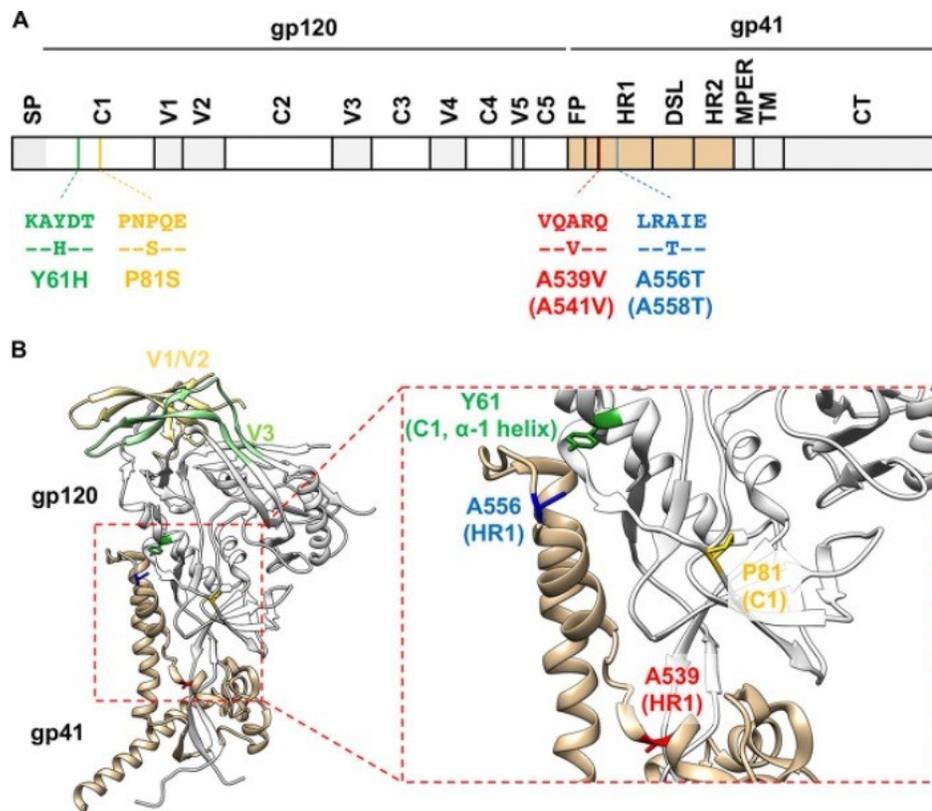


**Figure 4.** HIV-1 Replication cycle. The sites of FDA-approved antiretroviral therapies (ART) are indicated and numbered 1-4 in black (Kleinpeter & Freed, 2020).

There are currently five different classes of antiretroviral therapies (ART) which target various steps of the virus replication cycle (Arts & Hazuda, 2012). Entry inhibitors (Ent-Is) as seen in Figure 4, #1, prevent the fusion of the viral envelope and host cell membranes, so the

virus cannot enter the host cell (Kleinpeter & Freed, 2020). Nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs) (Figure 4, #2) prevent reverse transcriptase from reverse transcribing the viral RNA into complementary DNA (Kleinpeter & Freed, 2020). Integrase strand transfer inhibitors (INSTIs) (Figure 4, #3) target integrase preventing the viral DNA from being integrated into the host genome (Kleinpeter & Freed, 2020). Lastly, protease inhibitors (PIs) (Figure 4, #4) block the proteolytic cleavage of the *gag* and *gag-pol* polyprotein precursors producing immature virus particles (Kleinpeter & Freed, 2020).

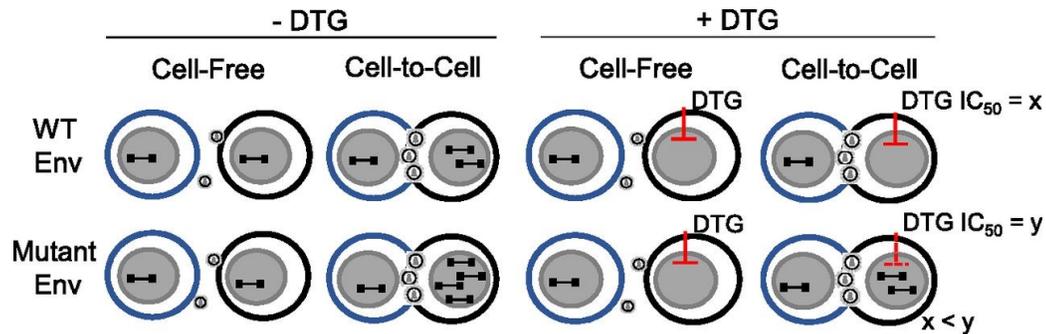
While ART has proven to be effective in preventing AIDS in people living with HIV (PLWH), resistance to antiretroviral drugs can emerge in individuals due to poor adherence and/or an insufficient drug regimen. HIV drug resistance tends to arise with the selection for or emergence of mutations in ART-target genes (Collier et al., 2019); however, virological failure can also occur in ART-treated individuals without mutations within target genes, particularly in those on PI and/or INSTI-containing regimens (Castain et al., 2019; Coetzer et al., 2017; Datir et al., 2020; Havlir et al., 2000; Hocqueloux et al., 2019; la Rosa et al., 2016; Manasa et al., 2017). These failures suggest that non-target gene mutations may contribute to drug resistance.



**Figure 5.** *Env* mutations, A539V and A556T were found to confer resistance to Dolutegravir (DTG; an integrase inhibitor) in vitro. (A) HIV-1 *env* gene diagram with the corresponding resistance mutations labeled in their appropriate subunit and domain in gp41. (B) *Env* structural model with the *env* mutations highlighted. Gp120 is colored white and Gp41 containing mutations is colored tan. Y61 mutation is shown in green, A556 mutation in blue, A539 mutation in red and P81 mutation in yellow. (Hikichi et al., 2021; Van Duyne et al., 2019)

In vitro studies have demonstrated that mutations in *env* can confer resistance to INSTIs (Van Duyne et al., 2019). In this study, HIV-1 virus with mutations in the p6 domain of *gag* exhibited severely impaired replication kinetics. However, when these p6-mutant viruses were propagated in cell culture, second-site compensatory mutations emerged in *env* in gp120 and gp41 (Figure 5). *Env* gp41 mutations, specifically A556T and A539V (Figure 5), rescued and enhanced viral replication kinetics but also displayed defects in cell-free particle infectivity. In addition, virus with *env* A539V or A556T mutations were shown to decrease the sensitivity of the virus to the INSTI, Dolutegravir (DTG), in vitro (Van Duyne et al., 2019). In the presence of DTG, A539V and A556T mutant viruses are thought to enhance spread via cell-cell transmission

resulting in an increased multiplicity of infection (MOI), circumventing the inhibitors (Figure 6) (Van Duyne et al., 2019). Cell-cell transmission occurs at virological synapses where *env* of the infected cell comes into contact with the CD4 receptor on the target cell (Figure 6) (Hikichi et al., 2021).



**Figure 6.** Model of cell-to-cell transmission of *env* mutant virus that have reduced sensitivity to DTG (Van Duyne et al., 2019).

Previously, we examined if these mutations within *env* are seen in vivo by performing *env* single-genome sequencing (SGS) on samples from participants failing a Raltegravir (RAL, an integrase inhibitor) containing regimen (Hikichi et al., 2021). SGS was also performed on *integrase* (*IN*) to examine the presence of any target gene mutations. *IN* sequencing revealed that only 1 of 5 participants had high levels (73%) of a known *IN* drug resistance mutation, N155H, whereas no *IN* mutations or low levels (<5%) were seen in the other 4 participants. *Env* sequencing identified many mutations (Table 1), some that were in multiple participants failing RAL (highlighted in yellow). Of note, all *env* mutations identified were seen at a higher frequency during early or late RAL failure compared to pre-RAL therapy. The frequency of residues at each of these positions was compared to 5,923 HIV-1 subtype C sequences from the Los Alamos HIV database and for several mutations, the amino acid mutation resulted in a change from a more conserved to a less conserved residue. *Env* mutations A607N, R633K and T641I were tested in vitro and found to confer resistance to DTG (Hikichi, et al., unpublished

data). Studies are on-going to evaluate several of these specific *env* mutations in vitro to determine their ability to decrease sensitivity to INSTIs. These findings suggest that these non-target gene mutations may contribute to the virological failure seen in these participants.

**A**

Participant identifiers	Mutations	Frequency (%)				Frequency of residues (%) in subtype C Env				
		Pre	Early	Late						
A1	gp120 C1	E32R <sup>a</sup>	4.8	0	ns	22	*	G: 51.32, V: 14.47, E: <b>9.21</b> , K: 6.58, N: 5.26, D: 3.95, S: 2.63, <b>R: 2.63</b> , other: 3.95		
		D62N <sup>a</sup>	0	0	ns	22	**		E: 75.53, <b>D: 18.54</b> , K: 2.34, <b>other: 3.58</b>	
		D113N <sup>a</sup>	7.1	44	*	22	ns		<b>D: 96.90</b> , <b>other: 3.10</b>	
	gp41	FP	V525A	4.5	0	ns	22		*	<b>A: 91.66</b> , <b>V: 3.31</b> , T: 3.10, other: 1.93
			HR2	N637S <sup>a</sup>	2.3	44	**		22	*
		R655K	6.8	0	ns	22	*		<b>K: 53.89</b> , Q: 21.09, <b>R: 10.82</b> , E: 7.17, N: 3.24, other: 3.79	
A3	gp120 C1	V87A	38	100	*	100	*	E: 56.04, G: 21.35, K: 10.33, <b>A: 3.35</b> <b>V: 3.28</b> , D: 1.74, other: 3.91		
	gp41	FP	T514G	3.6	80	**	75	**	<b>G: 95.63</b> , <b>other: 4.37</b>	
			M518V	7.1	80	**	100	***	<b>V: 77.67</b> , <b>M: 12.54</b> , L: 7.99, other: 1.79	
			A525V <sup>a</sup>	3.6	80	**	75	**	<b>A: 91.66</b> , <b>V: 3.31</b> , T: 3.10, other: 1.93	
	HR1	Q581L	32	100	**	100	*	<b>L: 99.24</b> , <b>other: 0.76</b>		
		V583I	3.6	80	**	75	**	<b>I: 78.08</b> , L: 14.06, <b>V: 5.31</b> , M: 2.55		
	DSL	V595M	18	80	*	75	*	I: 55.20, L: 36.04, <b>M: 7.65</b> , <b>other: 1.10</b>		
	HR2	K633R	32	100	**	100	*	<b>R: 73.19</b> , <b>K: 24.67</b> , other: 2.14		
		M644R	3.6	60	**	75	**	<b>R: 50.10</b> , K: 14.27, Q: 7.99, N: 6.20, S: 5.31, T: 3.45, E: 3.45, G: 2.96, W: 2.48, <b>other: 3.79</b>		
		E647A <sup>a</sup>	0	60	**	75	***	<b>E: 95.79</b> , <b>other: 4.21</b>		
A5	gp41	FP	T514G	26	N.D.	95	***	<b>G: 95.63</b> , <b>other: 4.37</b>		
			L518V	30	N.D.	95	***	<b>V: 77.67</b> , M: 12.54, <b>L: 7.99</b> , other: 1.79		
			F523L	30	N.D.	95	***	<b>L: 90.55</b> , <b>F: 7.72</b> , other: 1.72		
	HR1	A546S	30	N.D.	95	***	<b>S: 96.83</b> , <b>other: 3.17</b>			
	DSL	L602R <sup>a</sup>	26	N.D.	95	***	<b>L: 89.94</b> , I: 3.86, V: 1.65, <b>other: 4.55</b>			
	G624E	30	N.D.	95	***	D: 46.55, N: 20.07, <b>E: 14.90</b> , <b>G: 10.69</b> , K: 3.38, other: 4.41				
	HR2	A654T	33	N.D.	95	***	E: 97.59, <b>other: 2.41</b>			

## B

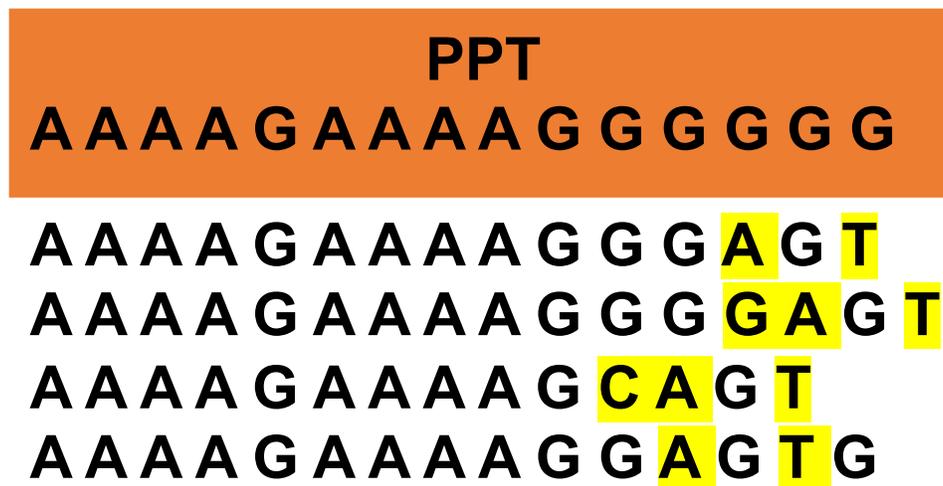
Participant identifiers	Mutations	Frequency (%)			Frequency of residues (%) in subtype C Env	
		Pre	Early	Late		
A4	gp120 C1	N57D	17	100 ***	100 ***	D: 93.52, N: 3.79, other: 2.69
		G60A	0	100 ***	100 ***	A: 89.32, G: 7.10, other: 3.58
		K63Q <sup>a</sup>	0	100 ***	100 ***	K: 50.17, R: 28.12, T: 15.02, Q: 1.86, other: 4.82
		I68V	0	100 ***	100 ***	V: 88.97, I: 9.93, other: 1.10
		D/N78S <sup>a</sup>	0	100 ***	100 ***	D: 98.14, other: 1.86
		I/L84M	33	100 *	100 *	M: 35.98, I: 35.08, L: 21.85, V: 5.03, other: 2.07
		M/K/I85V	26	100 *	100 *	V: 65.33, F: 8.20, E: 4.20, I: 4.14, L: 3.79, D: 2.34, A: 2.07, K: 1.93, P: 1.31, N: 1.17, R: 1.10, other: 4.41
		G/K87E	32	100 *	100 *	E: 56.04, G: 21.35, K: 10.33, A: 3.35, V: 3.28, D: 1.74, other: 3.91
		H105Q <sup>a</sup>	3.3	100 ***	100 ***	H: 95.31, other: 4.69
		T130N	3.3	100 ***	100 ***	N: 62.76, K: 7.24, E: 6.90, T: 4.76, H: 4.48, D: 4.21, R: 3.31, S: 2.55, other: 3.79
		gp41	FP	L515I	35	48 ns
FPPR	L529F			18	100 ***	100 ***
HR1	I535M <sup>a</sup>		0	41 ***	93 ***	I: 79.39, M: 8.61, V: 6.00, L: 5.72, other: 0.28
	A536T		35	59 ns	6.7 *	T: 89.32, A: 9.24, other: 1.45
HR1	S553N <sup>a</sup>		0	40.7 ***	93 ***	S: 83.87, N: 15.02, other: 1.10
	R557K <sup>a</sup>		0	59.3 ***	6.7 ns	R: 88.35, K: 10.48, other: 1.17
	A578T <sup>a</sup>		0	100 ***	100 ***	A: 59.68, T: 40.25, E: 0.07
	I583L <sup>a</sup>		0	100 ***	100 ***	I: 78.08, L: 14.06, V: 5.31, M: 2.55
DSL	L595I		0	100 ***	100 ***	I: 55.20, L: 36.04, M: 7.65, other: 1.10
	K601I <sup>a</sup>		0	100 ***	100 ***	K: 98.14, other: 1.86
	P605T		0	100 ***	100 ***	T: 95.24, other: 4.76
	A607N <sup>a</sup>		0	100 ***	100 ***	A: 51.96, N: 30.88, T: 10.06, S: 4.20, other: 2.89
	S612D <sup>a</sup>		0	100 ***	100 ***	S: 75.40, T: 6.20, A: 4.55, N: 4.00, I: 2.00, D: 1.93, V: 1.24, other: 4.69
	R617K		0	100 ***	100 ***	K: 84.13, R: 14.63, other: 1.24
	E620K <sup>a</sup>		0	100 ***	100 ***	E: 22.83, T: 19.79, D: 17.17, N: 7.52, K: 7.52, S: 6.97, A: 6.07, G: 4.97, Q: 2.97, other: 4.21
	Y621E		0	100 ***	100 ***	D: 53.27, E: 34.67, Y: 3.58, A: 2.55, Q: 2.14, other: 3.79
	G624D		0	100 ***	100 ***	D: 46.55, N: 20.07, E: 14.90, G: 10.69, K: 3.38, other: 4.41
	N625K <sup>a</sup>		0	100 ***	100 ***	N: 96.35, other: 3.65
HR2	R633K <sup>a</sup>		0	100 ***	100 ***	R: 73.19, K: 24.67, other: 2.14
	D636S		0	100 ***	100 ***	S: 53.55, N: 31.36, D: 11.58, other: 3.51
	T641I <sup>a</sup>		27	0 **	0 **	T: 72.78, I: 21.23, L: 1.65, other: 4.34
	N644G		0	100 ***	100 ***	R: 50.10, K: 14.27, Q: 7.99, N: 6.20, S: 5.31, T: 3.45, E: 3.45, G: 2.96, W: 2.48, other: 3.79
	I648E		0	100 ***	100 ***	E: 38.83, D: 38.69, V: 9.10, K: 5.38, I: 3.31, other: 4.69
	N651A/T <sup>a</sup>	0	100 ***	100 ***	N: 54.72, I: 16.26, T: 12.61, S: 11.30, K: 0.76, other: 4.34	
	K655N <sup>a</sup>	0	100 ***	100 ***	K: 53.89, Q: 21.09, R: 10.82, E: 7.17, N: 3.24, other: 3.79	
	E658K	0	100 ***	100 ***	K: 81.53, Q: 13.65, other: 4.82	
	K662A/T	12	100 ***	100 ***	A: 89.87, E: 7.51, other: 2.62	
	N665K	0	100 ***	100 ***	S: 66.09, K: 26.46, N: 3.24, other: 4.20	

**Table 1.** Previously identified *env* mutations in participants failing RAL: A) Participants A1 (P1), A3 (P3) and A5 (P5); B) Participant A4 (P4). Mutations highlighted in yellow indicate mutations identified in multiple participants. Frequency of residues at the corresponding position were compared to HIV-1 subtype C sequences from the Los Alamos HIV database (n=5,923). (Hikichi et al., 2021)

In addition to *env* mutations, mutations in *gag* and *nef* have also been examined in vitro to investigate their contribution to INSTI drug resistance. An *IN* mutation, E157Q was identified in a participant on DTG-containing therapy (Danion et al., 2015). The participant did not have virologic failure and in vitro this mutation did not affect resistance to RAL or DTG (Saladini et al., 2017). Hachiya, et al., and Sarafianos (2019) serially passed the E157Q mutant virus in the presence of DTG to evaluate whether other mutations may emerge that could lead to drug

resistance. They identified the G19S mutation in *gag* NC zinc finger, which in combination with the E157Q integrase mutation, demonstrated significant DTG resistance in vitro. The G19S nucleocapsid mutation is believed to affect reverse transcription causing changes in the proviral DNA ends. These changes are thought to affect DTG binding leading to a destabilized IN/vDNA/DTG complex, resulting in DTG resistance (Hachiya et al., 2019).

Mutations in the *nef* 3' polypurine tract (3' PPT) have also been identified in vitro to confer resistance to DTG with no resistance mutations in the *IN* target gene. Malet et al. (2017), performed serial passage of HIV-1 in the presence of DTG. No mutations were identified in *IN*, but 4 mutations were found in the 3' PPT of *nef*, shown in Figure 7. Additional studies are needed to examine if these mutations can be identified in vivo from participants failing INSTIs and to investigate to what extent that these mutations contribute to virologic failure.



**Figure 7.** 3' PPT region in HIV-1 *nef* is illustrated in orange. The nucleotides highlighted in yellow represent mutations identified in vitro in the 3' PPT region that confer resistance to DTG (Malet et al., 2017).

In our current study, we investigated whether mutations in non-target genes, *gag* Nucleocapsid (NC) and/or *nef* (3'PPT) might also contribute to drug resistance and virologic failure in five individuals with HIV failing RAL that were previously evaluated for *IN* and *env* mutations (Tables 1 and 2). To identify possible resistance mutations in *gag* or *nef* genes, we performed SGS on plasma-derived viruses from pre-RAL therapy and early and/or late RAL failure samples from the five participants. *Gag* and *nef* sequences were analyzed for mutations and phylogenetic trees were assembled with statistical analysis to determine evolutionary patterns of the viral sequences.

## **Materials and Methods**

### Participants

Samples were obtained from 5 participants from the AIDS Clinical Trial Group (ACTG) study A5273 which is a randomized, open-label phase-3 noninferiority study in nine resource-limited countries (three sites in India and South Africa, two in Malawi and Peru, and one in Brazil, Kenya, Tanzania, Thailand and Zimbabwe) (la Rosa et al., 2016). Each participant, following at least 24 weeks on non-nucleotide reverse transcriptase inhibitor (NRTI)-containing regimen and with HIV-1 RNA concentrations greater than 1000 copies/mL, were selected and received second-line therapy containing lopinavir/ritonavir (protease inhibitors) combined with raltegravir (RAL, integrase inhibitor). Plasma samples were obtained from a timepoint prior to RAL-containing regimen initiation and from 1-2 timepoints during increasing viremia due to failure of the RAL-containing regimen. At the time of sampling during RAL failure, each participant had acceptable therapeutic drug levels (RAL Conc > 33 ng/mL).

### Plasma viral RNA extraction

Plasma was centrifuged at 5,300xg for 10 min at 4°C to remove the cell debris. The supernatant was then spun at 16,000xg at 4°C for 1 h. The supernatant was discarded and 100µL of 3M GUHCL mix with fresh Protease K was added to the pellet. This tube was then placed in a water bath at 42°C for 1 h. The tube was then flash spun and 400µL of the GUSCN mix with fresh glycogen was added along with 500µL of room temperature 100% isopropanol. The tube was centrifuged at 21,000xg for 15 min at 23°C. The pellet was then washed with 750µL 70% ethanol and centrifuged at 21,000xg for 15 min at room temperature. The tube was spun again to separate the residual ethanol from the viral RNA to prevent the ethanol from inhibiting the PCR. The pellet was air dried until it turned translucent, then the RNA was resuspended in 20µL of 5mM TRIS-HCL (pH 8.0) (Palmer et al., 2005).

### cDNA synthesis

To synthesize cDNA from the extracted viral RNA, 2.5 µL of 10mM dNTPs and 2.5 µL of 50 µM oligo dT were added to the RNA sample in a 96 well PCR plate. The RNA was denatured in the thermocycler at 65°C for 10 min. A cocktail was made consisting of 5 µL of 10X RT buffer, 10 µL of 25 mM MgCl<sub>2</sub>, 0.5 µL of 0.1M Dithiothreitol (DTT), 8.5 µL of RNase-free water, 0.5 µL of Rnase-Out and 0.5 µL of 200 U/µL Superscript III Reverse Transcriptase. Once the RNA denaturation was complete, 25 µL of this cocktail was added to each sample. The cDNA was synthesized in the thermocycler for 50°C for 1 h, then 55°C for 1 h and then 70°C for 15 min. Once the cDNA program was complete, 2 µL of 2 U/µL RNaseH was added and the RNA was digested at 37°C for 20 min (Wiegand et al., 2017).

### Single-Genome Sequencing of *gag* and *nef*

To ensure that the *gag* and *nef* PCR products result from a single molecule of plasma HIV RNA, serial dilutions of the cDNA were prepared, and the endpoint was determined where approximately  $\leq 30\%$  of the wells are positive. A PCR master mix was made with 100  $\mu\text{L}$  of 10X PCR buffer, 40  $\mu\text{L}$  of 50 mM  $\text{MgCl}_2$ , 20  $\mu\text{L}$  of 10mM dNTPs, 8  $\mu\text{L}$  of Platinum Taq Enzyme, 624  $\mu\text{L}$  of molecular-grade water and 4  $\mu\text{L}$  of each 50  $\mu\text{M}$  primer. The primers used for the first round of PCR for *gag* were Gag.OF (5'-TAAAGCTTGCTTGAGTGC-3') and Gag.OR (5'-TACTGTATCATCTGCYCCTG-3'). The PCR cycling conditions for the first round of *gag* PCR reactions was performed as follows: 94°C for 2 min, 44 cycles 94°C for 30 s, 53°C for 30 s and 72°C for 2 min, followed by a final elongation step at 72°C for 3 min. Nested PCR was then performed using the following primers, G00.IF (5'-GACTAGCGGAGGCTAGAAG-3') and G01.IR (5'-AGGGGTCGTTGCCAAAGA-3') (Sanders-Buell et al., 1995). The PCR cycling conditions were performed as follows: 94°C for 2 min, 44 cycles 94°C for 30 s, 57°C for 30 s and 72°C for 1 min 30 s, followed by a final elongation step at 72°C for 3 min.

To acquire *nef* PCR products from a single cDNA molecule, a single round of *env-nef* PCR was performed. The specific cDNA endpoint dilution was used with the following primers (modified from Salazar-Gonzalez et al., 2008) E0 FWD (5'-TAGAGCCCTGGAGCATCCAGGAAGTCAGCCTA-3') (Sanders-Buell et al., 1995) and HIVC.short.OFM.R1 (5'-CAAGGCAAGCTTTATTGAGGCTTA-3') (Salazar-Gonzalez et al., 2008). The following PCR cycling conditions were then performed: 95°C for 2 min and 44 cycles of 95°C for 30 s, 57°C for 30 s, and 68°C for 3 min to 3 min 30 s, with a final elongation step at 68°C for 5 min. Nested PCR was performed using the following primers: 8351F (5'-ATAGAGTTAGGCAGGGATACTCACC-3') and HIVC.short OFM.R1 (5'-

CAAGGCAAGCTTTATTAGGCTTA- 3'). The following PCR cycling conditions were then performed: 94°C for 2 min, 44 cycles of 94°C for 30s, 57°C for 30s and 72°C for 1min, followed by a final elongation of 72°C for 3 min.

To identify positive wells, detection using Gel RED read-out plates was performed. In a 15mL falcon tube, 10mL of 5mM TRIS-HCL and 4µL of Gel RED were combined and 10µL was added to each well of a 96 well plate. The PCR product was diluted by adding 60µL of 5mM TRIS-HCl and 10µL of the diluted PCR product was added to each well with the Gel RED. The Gel RED read-out plate was exposed to UV light and an image was acquired. Bright wells in the image correspond to positive PCR product in the wells. The positive samples were sequenced by Sanger. For *gag* sequencing, G00.IF and G01.IR nested primers and the following primers were used: 1893R (5'-CTCATTGCCTCAGCCAAAACCTC-3') and P24-1.OF (5'-AGYCAAAAATTAYCCYATAGT-3') (Swanson et al., 2004). For *nef* sequencing, HIVC.short.OFM.R1 *env* primer and the following primers were used: 8351F (5'-ATAGAGTTAGGCAGGGATACTCACC-3'), 8538F and HIVCshort.ENVNR2 (5'-CAATCAGGGAAGTAGCCTTG-3').

### Sequence Analysis

Sanger sequencing data was analyzed using MEGA7 ([www.megasoftware.net](http://www.megasoftware.net)) to align sequences and build neighbor-joining phylogenetic trees. *Gag* sequences were analyzed for mutations within the Nucleocapsid (NC) and *nef* sequences were analyzed for mutations in the 3' polypurine tract (PPT).

### Statistical Analyses

Average pairwise distances (APD) were measured to determine if there was increasing or decreasing genetic diversity over time. Using MEGA7 ([www.megasoftware.net](http://www.megasoftware.net)), APD were calculated using the p-distance method. Panmixia (measure of divergence) was analyzed using an in-house program and a significance cutoff of less than  $10^{-3}$  applied. This analysis measures whether there is a genetic shift in the population from pre-RAL to RAL failure by analyzing the divergence in populations over time. For populations to be considered statistically different, indicating a genetic shift in the population, the probability of panmixia is less than 0.001 (Kearney et al., 2014). Otherwise, a population is considered well-mixed or “panmictic”, meaning there is no discernible segregation between them, indicating that the population has not shifted genetically. Root-to-tip distances were plotted against the time of sample collection (pre-ART versus virological failure) and evaluated via F-test on the resulting linear regression. Root-to-tip analysis determines if changes may be due to the emergence of new variants or to the selection of variants that were already present within the population prior to ART. Units for slopes for these analyses are substitutions/month.

### **Results**

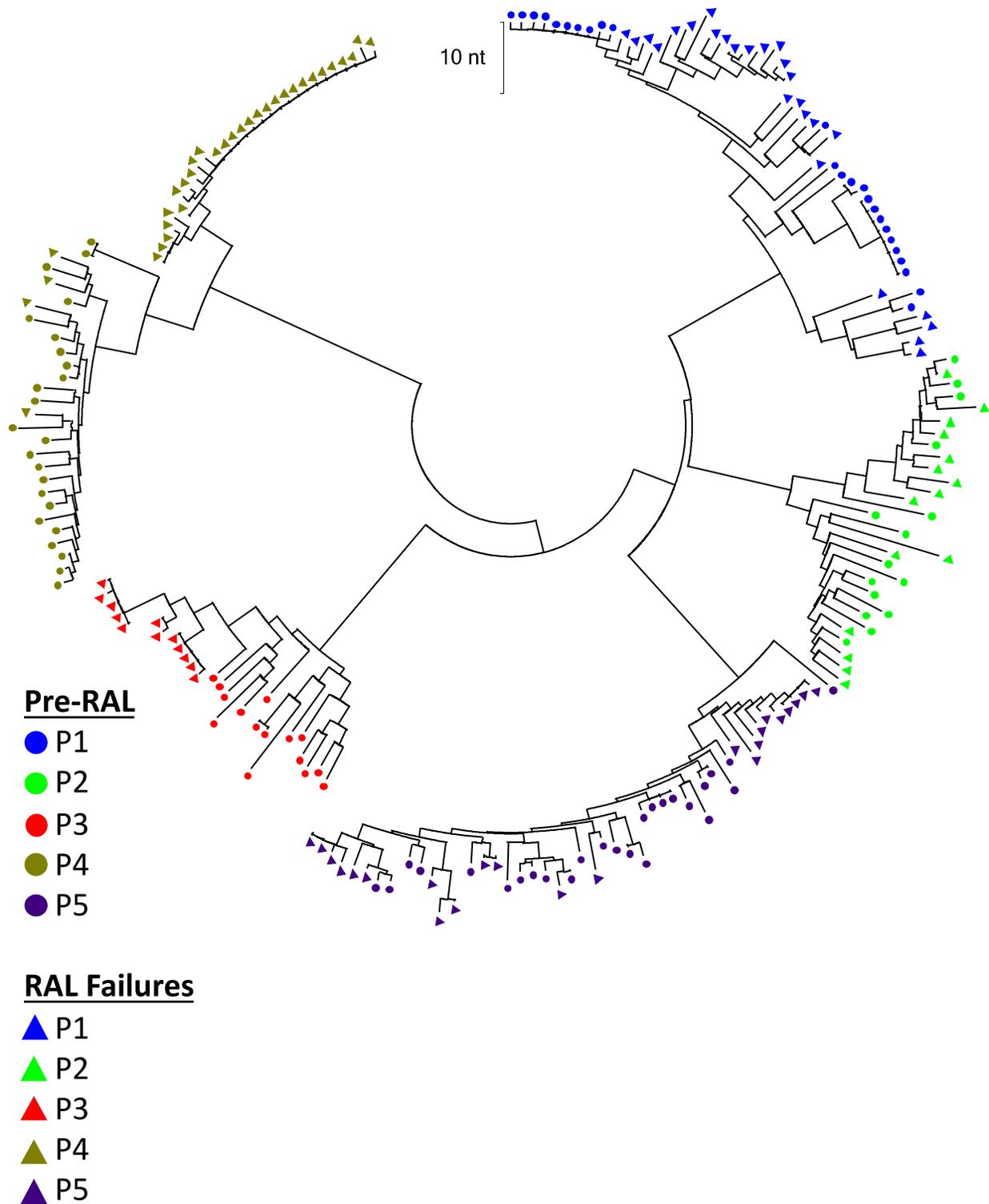
Participant samples in this study were from the ACTG A5273 SELECT study, which evaluated whether a boosted lopinavir plus RAL was non-inferior to a boosted lopinavir plus NRTIs by assessing their ability to clinically suppress HIV-1 infection and prevent virologic failure in resource-limited setting. We obtained plasma samples from 5 participants prior to initiation of RAL-containing therapy and from 1-2 timepoints during ART failure, illustrated in Table 2. Plasma HIV-1 RNA levels ranged from 30,206 - 1,617,523 copies/ml in Pre-RAL

therapy, 433-7200 copies/ml in early RAL-failure and 1139-11,335 copies/ml in late RAL-failure samples. At the time of sample collection during RAL-failure, all the participants had acceptable levels of RAL detected (RAL Conc. > 33 ng/mL) in their plasma. Samples from all 5 participants were previously examined for target mutations in *IN* (Table 2) and non-target mutations in *env* by SGS (Table 1). No known *integrase* resistance mutations were identified in any of the participants prior to initiation of RAL therapy. However, during early and late RAL failure, *integrase* sequences from participant P4 had a high frequency (73%) of the N155H *IN* drug resistance mutation. All other participant sequences from early and late RAL failure, had only low frequency of *integrase* resistance mutations (Table 2). The low frequency of *integrase* resistance mutations suggests that mutations in non-target genes, such as *env* identified in our previous studies, may contribute to RAL failure. Here, we examined whether other non-target mutations in *gag* NC and *nef* 3'PPT described from in vitro studies (Hachiya et al., and Sarafianos, 2019; Malet et al., 2017) could also contribute to INSTI resistance in these participants. Additionally, sequences were analyzed for genetic diversity, divergence, and whether sequences identified during RAL-failure were the result of the evolution of new variants or selection for variants already present in the population prior to ART for each participant.

PID	Sampling Timepoint		HIV-1 RNA (copies/ml)	RAL conc. (ng/ml)	IN resistance mutations
<b>P1</b>	Pre-ART	Day 0	805,710	-	No mutations detected
	Early RAL failure	Day 181	570	990	Not determined
	Late RAL failure	Day 250	11,335	89	<b>N155H (5.6%)</b>
<b>P2</b>	Pre-ART	Day 0	17,450	-	No mutations detected
	Late RAL failure	Day 195	1,772	93	No mutations detected
<b>P3</b>	Pre-ART	Day 0	1,329,415	-	No mutations detected
	Early RAL failure	Day 178	433	965	Not determined
	Late RAL failure	Day 196	1,139	3,371	<b>F121Y (14%)</b>
<b>P4</b>	Pre-ART	Day 0	1,617,523	-	No mutations detected
	Early RAL failure	Day 175	7,208	1,142	<b>E138K (2.7%), N155H (73%)</b>
	Late RAL failure	Day 199	9,918	480	Not determined
<b>P5</b>	Pre-ART	Day 0	30,206	-	No mutations detected
	Late RAL failure	Day 194	18,806	2,729	<b>G163K (4.8%)</b>

**Table 2.** Participant sample information

To explore this possibility, we performed SGS of *gag* and *nef* in plasma viral RNA from the 5 participants and compared sequences obtained from pre-RAL therapy to RAL failure to identify any potential RAL resistance mutations. For *gag* SGS, a 1488nt fragment was amplified and sequenced. To verify that all *gag* sequences obtained and analyzed cluster to their respective participant sample, we performed a quality control check by building a phylogenetic tree containing all participant pre-RAL and RAL failure sequences (Figure 8). As shown in Figure 8, all *gag* sequences cluster to their respective participant indicating that there was no cross contamination or mislabeling of the *gag* sequences.



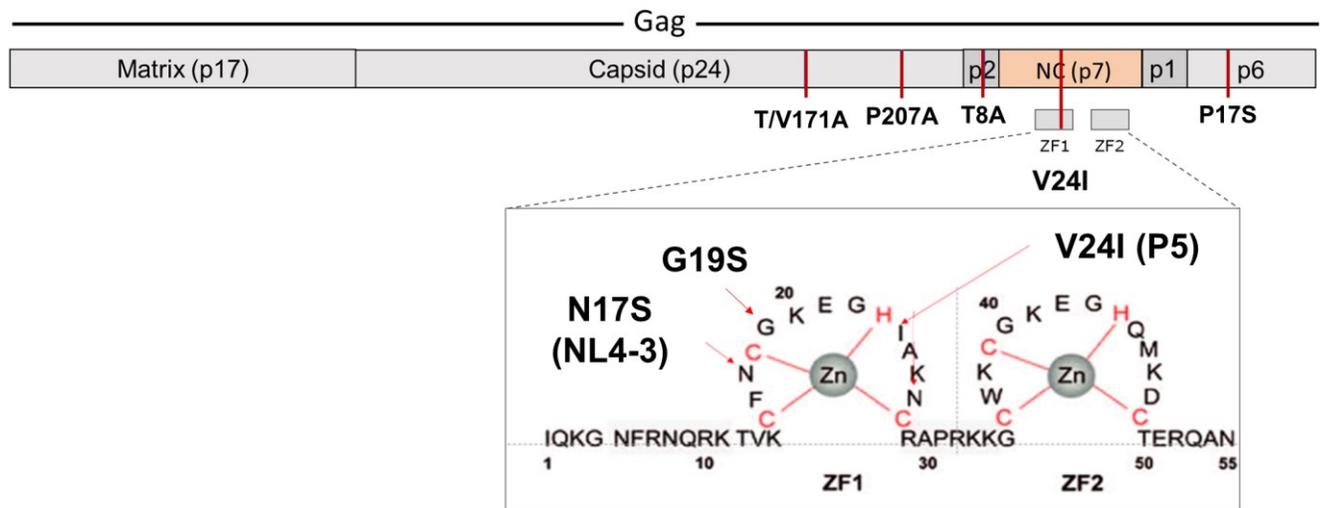
**Figure 8.** Phylogenetic tree of all five participant's pre-RAL and RAL failure *gag* sequences for quality control purposes.

To identify specific mutations within *gag* and particularly in NC, we compared pre-RAL and late RAL failure sequences for each patient. Previous in vitro studies identified viruses with the NC-G19S mutation in conjunction with *IN-E157Q* (Hachiya et al., 2019) and NC-N17S (Hikichi & Freed, unpublished data), which showed a decrease in sensitivity of the virus to DTG. Table 3 lists the mutations found in NC from the participants in this study. No NC mutations were identified for P1 or P4, however, in the remaining 3 participants, we found NC mutations that increased in frequency during RAL failure compared to pre-RAL. In pre-RAL, 2 different amino acids were present at a specific position across sequences but in RAL failure, one amino acid seemed to be selected and increase in frequency across sequences (i.e., Q/K41K; pre-RAL: 27% and RAL failure: 60%). For P5, the V24I mutation changes the amino acid residue from a higher frequency to lower frequency residue when compared to amino acids in that position from 2129 HIV-1 NC sequences in the Los Alamos HIV database.

Participant Identifiers	Mutations	Frequency (%)		Frequency of residues (%) in subtype C Nucleocapsid
		Pre-RAL	Late RAL Failure	
1	n/a	-	-	-
2	R/K10K	27	47	K: 61.5, R: 35.6, other: 0.6
	Q/K41K	27	60	K: 83.5, R: 8.5, Q: 5.0, other: 1.0
3	R/K3K	20	100	R: 76.4, K: 21.4, other: 1.1
	G/S4S	40	100	S: 61.4, G: 21.0, N: 14.7, other: 0.2
	K/R10R	20	100	K: 61.5, R: 35.6, other: 0.6
	I/V13I	53	100	V: 65.6, I: 28.8, other: 0.9
4	n/a	-	-	-
5	V24I	12	43	V: 65.6, I: 28.8, other: 0.9

**Table 3.** Frequency of the specific mutations observed in Nucleocapsid region of *gag* in the pre-RAL and RAL failure sequences for P1-P5. Frequency of residues in subtype C NC (n=2129) were determined using the Los Alamos HIV database.

We further evaluated the V24I mutation in NC from P5 and illustrated its location in Figure 9. This mutation is in the zinc finger domain and is adjacent to the amino acid that binds to zinc, similar to the previously discussed G19S (Hachiya et al., 2019) and N17S (Hikichi & Freed, unpublished data) mutations that are associated with decreased sensitivity of the virus to DTG. P2 also had a mutation within the zinc finger domain, Q/K41K, but this mutation was not adjacent to an amino acid binding zinc. Further in vitro studies are necessary to evaluate whether these zinc finger mutations or any other NC mutations are associated with INSTI drug resistance.



**Figure 9.** Nucleocapsid (NC) mutation identified from a participant sample during RAL failure and from in vitro studies localizes next to amino acids that associate with metal ion in a zinc finger domain.

We also evaluated the *gag* sequences from all 5 participants for mutations in the gene outside of the NC region. Table 4 lists *gag* mutations that were identified per participant along with the frequency of the mutation from pre-RAL to early/late RAL failure. All the mutations increased in frequency during RAL failure, but the significance of these mutations is unknown. Further analysis of these mutations and testing in vitro is needed to determine the possibility of their role in conferring INSTI drug resistance.

Participant Identifiers	Mutations	Frequency (%)		
		Pre-RAL	Early Failure	Late Failure
1	H219Q	0	-	40
	D/E55E	33	-	60
2	I/A60A	33	-	60
	T/Q90Q	33	-	60
	K/I92I	33	-	60
	R/L95L	33	-	60
	R/G123G	33	-	60
	E/V125V	33	-	60
	V/I129V	60	-	80
	V/I439I	27	-	60
	S/P442P	27	-	60
	H/Y443Y	27	-	60
3	K/H28K	60	-	100
	I/L31I	60	-	100
	V35I	0	-	42
	T53S	0	-	42
	V/T84V	73	-	100
	D/G125G	67	-	100
	A/T214A	67	-	100
	V/G377V	67	-	100
	T/N476A	0	-	100
A/S496S	40	-	100	
4	G91N	13	100	84
	T117A	0	64	74
	K119Q	0	100	80
5	T/V305A	4	-	39
	P341A	0	-	39
	T373A	4	-	39
	P468S	15	-	48

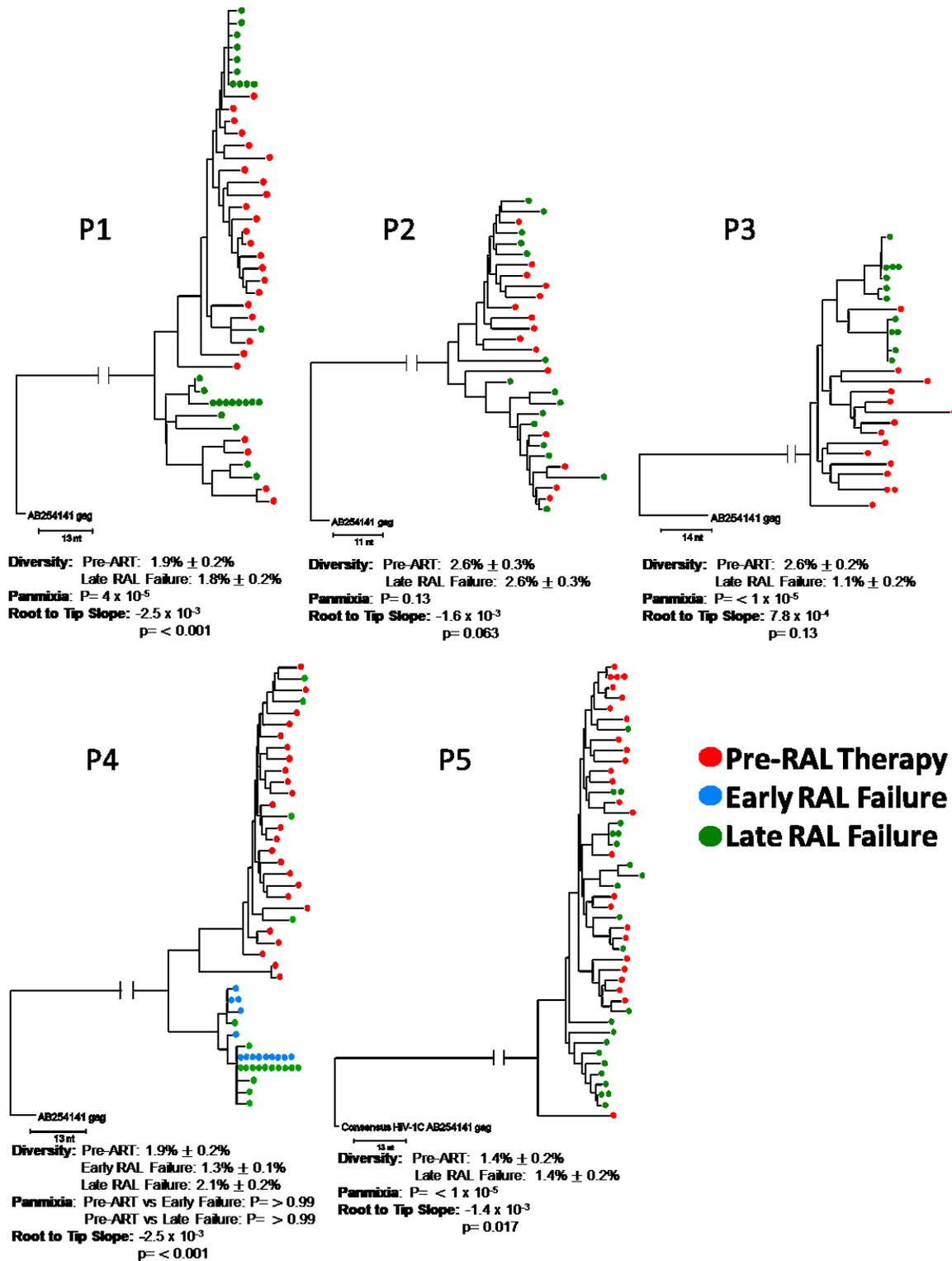
**Table 4.** Mutations found in other regions of *gag* from the 5 participants of this study along with the frequencies of the mutations in the pre-RAL and failure sequences.

Individual neighbor-joining phylogenetic trees for each participant were also constructed from *gag* sequences and are illustrated in Figure 10. The average pairwise distance (APD) (measurement of genetic diversity) was calculated for each sample timepoint: pre-RAL therapy and early and/or late RAL failure for each participant (Figure 10) to determine whether there was an increase/decrease or no change in the genetic diversity during RAL failure. In 4 participants (P1, P2, P4 and P5), there was no change in genetic diversity from pre-RAL to late RAL failure. For P4, there was a slight decrease in diversity in the viral population from pre-RAL to early RAL failure, 1.9% to 1.3%; however, diversity during late RAL failure (2.1%) was not different compared to pre-RAL sequences. P3 showed a decrease in genetic diversity from 2.6% to 1.1%, suggesting a genetic bottleneck with RAL failure.

To determine whether there was a genetic shift in *gag* during RAL failure, we calculated panmixia, which measures the significance of the change in population over time (divergence). For example, if sequences from pre-RAL and RAL failure can randomly mate with either population (“look like each other”) or are not statistically different, they are described as “panmictic”, meaning that the populations are well-mixed. However, if the populations are statistically different from each other with a probability of panmixia  $<0.001$ , then the populations are divergent and, therefore, diverged during failure. A genetic shift in population during RAL failure was seen in P1, P3 and P5 and illustrated by segregation of RAL failure sequences from pre-RAL therapy sequences. For P2 and P4, no apparent genetic shift in population was observed and pre-RAL and RAL failure samples are intermixed on the respective trees.

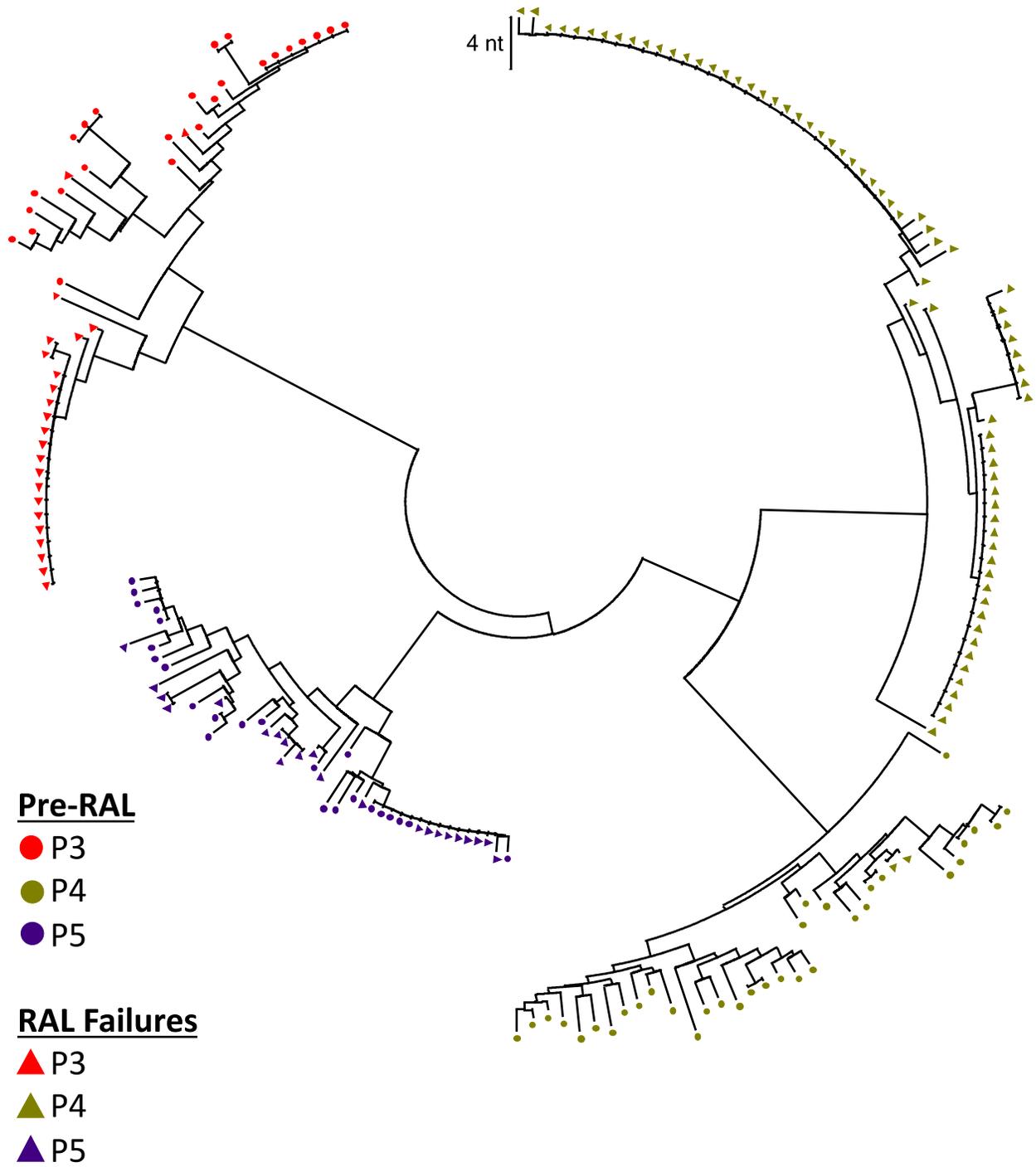
To determine if a genetic shift in population due to RAL failure is due to the emergence of new variants or to the selection for variants present in pre-RAL population, root-to-tip distances were measured using maximum likelihood trees and slopes are calculated by linear

regression. The units for the slopes for these analyses are substitutions per month. On trees with clock-like structure (look like a “staircase”), positive slopes suggest emergence of new variants and negative or no slopes suggest that variants were present in the pre-RAL population. However, the interpretation of trees with separately branched populations may be more complex. Only P3 had a positive root-to-tip slope ( $7.8 \times 10^{-4}$ ) and RAL failure sequences appeared separately branched from pre-RAL sequences, suggesting that accumulation of mutations during RAL failure may be newly emergent and not in the pre-existing population. For P1, P2, P4 and P5, the root-to-tip slopes were all negative ( $-2.5 \times 10^{-3}$ ,  $-1.6 \times 10^{-3}$ ,  $-2.5 \times 10^{-3}$  and  $-1.4 \times 10^{-3}$ , respectively), which suggests that variants during RAL failure were likely present prior to treatment.



**Figure 10.** Phylogenetic trees of *gag* in P1-P5 and the participant's corresponding diversity and panmixia values for pre-RAL and RAL failure.

Previous in vitro studies (Malet et al., 2017) found that mutations in *nef*, specifically in the 3' PPT region, may be responsible for drug resistance in individuals with HIV on ART. In this study, we performed *nef* SGS on samples from three out of five participants to identify mutations in *nef* during RAL failure. For *nef* SGS, a 1314nt fragment was amplified and sequenced. To verify that all *nef* sequences obtained and analyzed cluster to their respective participant sample, we performed a quality control check by building a phylogenetic tree containing all the participant pre-RAL and RAL failure sequences (Figure 11). As shown in Figure 11, all *nef* sequences cluster to their respective participant indicating that there was no cross contamination or mislabeling of the *nef* sequencing.



**Figure 11.** Phylogenetic tree of pre-RAL and RAL failure *nef* sequences from 3 participants for quality control purposes.

To identify specific mutations within *nef* and particularly in 3'PPT region, we compared pre-RAL and RAL failure sequences for each participant. Previous in vitro experiments (Malet et al., 2017) have found mutations in the 3'PPT region of *nef* that have been associated with INSTI resistance (Figure 7). We did not identify any 3'PPT mutations in any *nef* sequences from the participant samples. However, we found mutations in *nef* outside of the 3'PPT region that encoded amino acid changes in P3, P4 and P5, which are listed in Table 5. All *nef* mutations increased in frequency during RAL failure compared to pre-RAL samples. The significance of these mutations is unknown and further analysis and testing in vitro is required to assess their possible contribution to the development of INSTI drug resistance.

Participant Identifiers	Mutations	Frequency (%)		
		Pre-RAL	Early Failure	Late Failure
3	R9S	0	80	94
	A15T	0	80	89
	E18D	0	80	94
	R/K24E	0	80	94
4	N3G	0	97	N.D.
	E18D	0	97	N.D.
	I20M	0	97	N.D.
	R21K	0	97	N.D.
	E34D	0	96	N.D.
	R45Q	0	97	N.D.
	E48D	0	97	N.D.
	R49K	6	97	N.D.
	H50Y	0	97	N.D.
	Q60T	0	97	N.D.
	A63E	0	97	N.D.
E76G	0	97	N.D.	
5	n/a	-	-	-

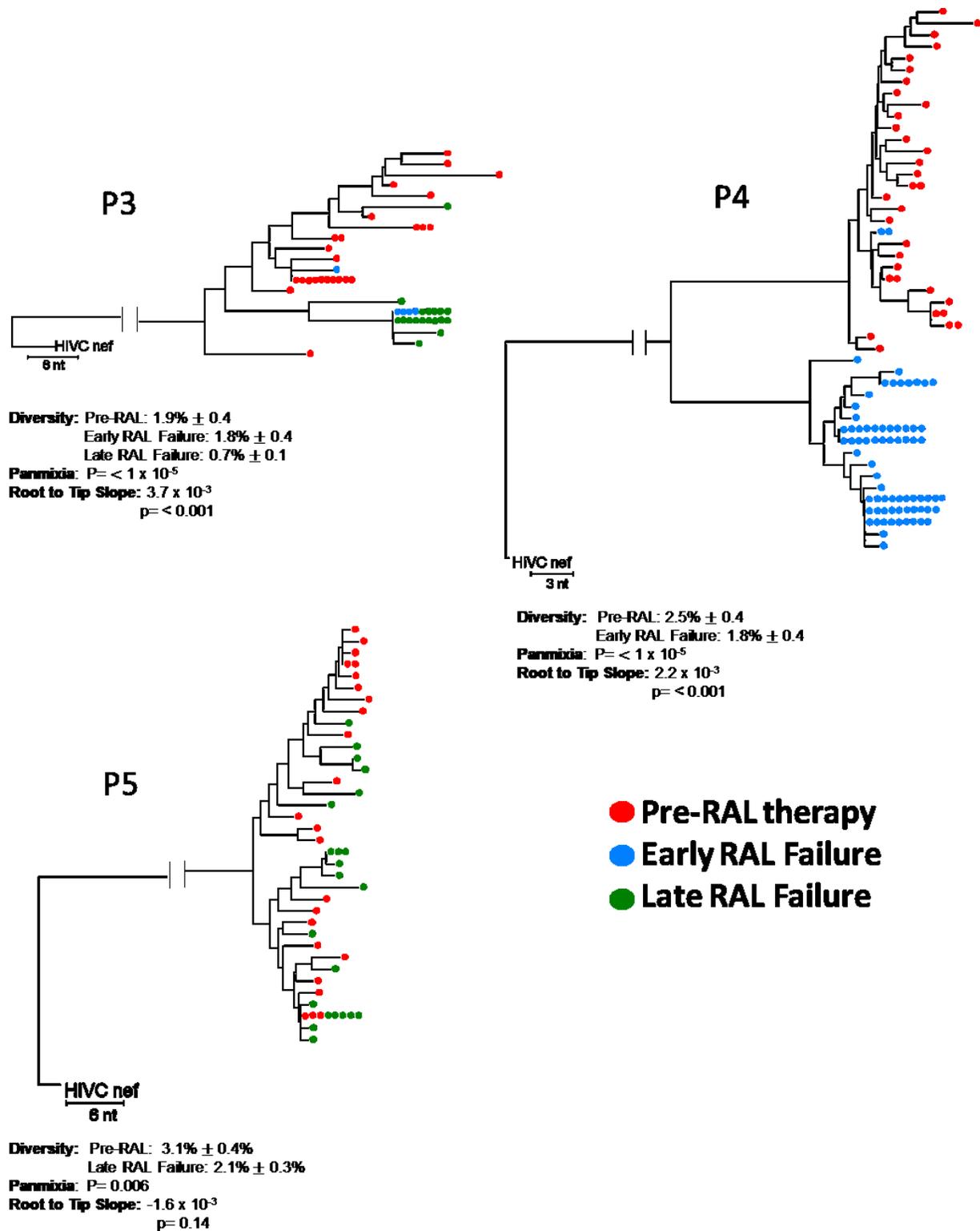
**Table 5.** *Nef* mutations found in P3, P4 and P5 of this study along with the frequencies of the mutations in the pre-RAL and failure sequences. N.D.= Sequences not done.

Individual neighbor-joining phylogenetic trees for each participant were constructed from the *nef* sequences and are illustrated in Figure 12. Like for the *gag* sequences, the APD were calculated for each sample timepoint (pre-RAL therapy and early and/or late RAL failure for each participant) (Figure 12) to determine if there was an increase/decrease or no change in the genetic diversity during RAL failure. For participant P3, there was no change in genetic diversity

from pre-RAL to early RAL failure, but there was a decrease in diversity in the viral population from pre-RAL to late RAL failure, 1.9% to 0.7%, suggesting a genetic bottleneck during late RAL failure. A decrease in diversity was seen for P4 from pre-RAL to early RAL failure, 2.5% to 1.8%, and for P5 from pre-RAL to late RAL failure, 3.1% to 2.1%, suggesting a genetic bottleneck during RAL failure for both participants.

To determine if there was a genetic shift during RAL failure, we calculated panmixia from participant-acquired *nef* sequences. For P5, no apparent genetic shift in the viral population was observed and pre-RAL and RAL failure samples are intermixed on the respective phylogenetic trees. A genetic shift in the population during RAL failure was seen in P3 and P4 and is illustrated by segregation of early RAL failure sequences from pre-RAL therapy sequences on phylogenetic trees.

To determine if the genetic shifts in the viral populations during RAL failure were due to the emergence of new variants or to the selection for variants present in the pre-RAL population, root-to-tip distances were measured using maximum likelihood trees and branch lengths were plotted. Slopes (substitutions per month) were calculated by linear regression. For participants, P3 and P4, the root-to-tip slopes were positive ( $3.7 \times 10^{-3}$  and  $2.2 \times 10^{-3}$ , respectively), which suggests that new variants may have emerged during RAL failure. Participant P5 had a negative root-to-tip slope ( $-1.6 \times 10^{-3}$ ), suggesting that mutations during RAL failure were likely present prior to treatment.



**Figure12.** Phylogenetic trees of *nef* in P3, P4 and P5 and the participant's corresponding diversity and panmixia values for pre-RAL and RAL failure.

## Discussion

In this study, we conducted SGS of *gag* and *nef* genes in plasma viral RNA from 5 individuals with HIV on a RAL-containing regimen and who experienced virologic failure within 175-250 days following therapy initiation. Sequences from 1 participant, P4, had a high frequency of known INSTI drug resistance mutations in the *integrase* gene explaining their ART failure. The remaining 4 participants had low levels of INSTI resistance mutations, suggesting that mutations in non-target genes may have contributed to their ART failure. In addition to the previously identified *env* mutations in these individuals, we found many mutations in the NC protein encoded by the *gag* gene. Most mutations found in NC increased in frequency during RAL failure. Specific amino acid residues in NC were mutated to a more conserved residue. However, we did identify three mutations in NC, K/R10K, I/V13I and V24I, that decreased in frequency. The NC-V24I mutation, located adjacent to an amino acid that binds to zinc in a zinc finger domain, is of interest since G19S and N17S are in a similar location and have been shown to confer resistance to DTG in previous in vitro studies (Hachiya et al., 2019; Hikichi & Freed, unpublished data). The working hypothesis for the G19S mutation is that it affects reverse transcription by changing the ends of proviral DNA (Hachiya et al., 2019). Thus, this change affects DTG binding to integrase since the IN/vDNA/DTG complex is destabilized, resulting in DTG resistance. To further explore whether the NC-V24I mutation contributes to a decrease in sensitivity to INSTIs in a similar mechanism, in vitro studies using a mutated virus would be necessary.

*Nef* was also sequenced in this study and analyzed for resistance mutations. Previous studies have determined that certain mutations in the 3'PPT of *nef* confer resistance to INSTIs in vitro (Malet et al., 2017). The mechanism is unknown; however, it is hypothesized that the

change in the 3'PPT could modify reverse transcription resulting in disrupted DTG binding (Malet et al., 2017). In our study, we examined *nef* sequences from participants with HIV failing RAL to identify 3'PPT *nef* mutations. No mutations were identified within the 3'PPT region, however, Table 5 shows mutations identified in other regions of *nef*. All the *nef* mutations identified increased in frequency during RAL failure. Currently, we are unsure of the significance of these mutations and further studies need to be done to determine if these mutations contribute to RAL drug resistance.

In future studies, we plan to continue *nef* SGS on samples from participants P1 and P2 to identify mutations that may contribute to RAL failure. To understand the significance of these mutations and their possible contribution to INSTI drug resistance, mutations will need to be cloned into a subtype C virus and evaluated in replication and infectivity studies in the presence of INSTIs. To determine if mutations present during RAL failure evolved in response to ART or were present prior to treatment and selected by ART, ultrasensitive single-genome sequencing could also be performed on specific gene regions. A limitation of this study is that some of the participants have smaller numbers of pre-ART and RAL failure samples, thus more samples, if available, may be needed to further evaluate RAL resistance mutations in these non-target genes.

In conclusion, the findings of this study demonstrate that mutations in off-target genes may contribute to the emergence of drug resistance in vivo. These findings indicate that it may be necessary to incorporate off-target genes in standard clinical genotyping protocols in the future to identify why some individuals with HIV fail ART without resistance mutations in the target genes.

## References

- Arts, E. J., & Hazuda, D. J. (2012). HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harbor Perspectives in Medicine*, 2(4), a007161. <https://doi.org/10.1101/CSHPERSPECT.A007161>
- Castain, L., Perrier, M., Charpentier, C., Palich, R., Desire, N., Wirden, M., Descamps, D., Sayon, S., Landman, R., Valantin, M.-A., Joly, V., Peytavin, G., Yazdanpanah, Y., Katlama, C., Calvez, V., Marcelin, A.-G., & Todesco, E. (2019). New mechanisms of resistance in virological failure to protease inhibitors: selection of non-described protease, Gag and Gp41 mutations. *Journal of Antimicrobial Chemotherapy* . <https://doi.org/10.1093/jac/dkz151>
- Cervera, L., Gòdia, F., Tarrés-Freixas, F., Aguilar-Gurrieri, C., Carrillo, J., Blanco, J., & Gutiérrez-Granados, S. (2019). Production of HIV-1-based virus-like particles for vaccination: achievements and limits. *Applied Microbiology and Biotechnology* 2019 103:18, 103(18), 7367–7384. <https://doi.org/10.1007/S00253-019-10038-3>
- Coetzer, M., Ledingham, L., Diero, L., Kemboi, E., Orido, M., & Kantor, R. (2017). Gp41 and Gag amino acids linked to HIV-1 protease inhibitor-based second-line failure in HIV-1 subtype A from Western Kenya. *Journal of the International AIDS Society*, 20(3). <https://doi.org/10.1002/JIA2.25024>
- Collier, D. A., Monit, C., & Gupta, R. K. (2019). *Cell Host & Microbe The Impact of HIV-1 Drug Escape on the Global Treatment Landscape*. <https://doi.org/10.1016/j.chom.2019.06.010>

- Danion, F., Belissa, E., Peytavin, G., Thierry, E., Lanternier, F., Scemla, A., Lortholary, O., Delelis, O., Avettand-Fenoel, V., & Duvivier, C. (2015). Non-virological response to a dolutegravir-containing regimen in a patient harbouring a E157Q-mutated virus in the integrase region. *The Journal of Antimicrobial Chemotherapy*, *70*(6), 1921–1923.  
<https://doi.org/10.1093/JAC/DKV012>
- Datir, R., Kemp, S., el Bouzidi, K., Mlchocova, P., Goldstein, R., Breuer, J., Towers, G. J., Jolly, C., Quiñones-Mateu, M. E., Dakum, P. S., Ndembu, N., & Gupta, R. K. (2020). In Vivo Emergence of a Novel Protease Inhibitor Resistance Signature in HIV-1 Matrix. *MBio*, *11*(6), 1–15. <https://doi.org/10.1128/MBIO.02036-20>
- Deeks, S. G., Overbaugh, J., Phillips, A., & Buchbinder, S. (2015). HIV infection. *Nature Reviews Disease Primers*, *1*. <https://doi.org/10.1038/nrdp.2015.35>
- Finzi, D., Blankson, J., Siliciano, J. D., Margolick, J. B., Chadwick, K., Pierson, T., Smith, K., Lisziewicz, J., Lori, F., Flexner, C., Quinn, T. C., Chaisson, R. E., Rosenberg, E., Walker, B., Gange, S., Gallant, J., & Siliciano, R. F. (1999). Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nature Medicine* *1999* *5*:5, *5*(5), 512–517.  
<https://doi.org/10.1038/8394>
- Finzi, D., Hermankova, M., Pierson, T., Carruth, L. M., Buck, C., Chaisson, R. E., Quinn, T. C., Chadwick, K., Margolick, J., Brookmeyer, R., Gallant, J., Markowitz, M., Ho, D. D., Richman, D. D., & Siliciano, R. F. (1997). Identification of a Reservoir for HIV-1 in Patients on Highly Active Antiretroviral Therapy. *Science*, *278*(5341), 1295–1300.  
<https://doi.org/10.1126/SCIENCE.278.5341.1295>

- Flint, J., Racaniello, V., Rall, G., Hatziioannou, T., & Skalka, A. (2020). *Principles of Virology* (5th ed.). ASM Press.
- Foley, B. T., Korber, B., Tina, M., Leitner, T. K., Apetrei, C., Hahn, B., Mizrachi, I., Mullins, J., Rambaut, A., & Wolinsky, S. (2018). *Title: HIV Sequence Compendium 2018 Intended for*. <https://www.hiv.lanl.gov/>
- Göttlinger, H. G. (2002). *HIV-1 Gag 2 Reviews HIV-1 Gag: a Molecular Machine Driving Viral Particle Assembly and Release*.
- Hachiya, A., Kirby, K., & Ode, H. (2019). Disruption of HIV-1 LTR sequence by a nucleocapsid mutation leads to DTG resistance. *Conference on Retroviruses and Opportunistic Infections, Seattle, Washington* .
- Havlir, D. v., Hellmann, N. S., Petropoulos, C. J., Whitcomb, J. M., Collier, A. C., Hirsch, M. S., Tebas, P., Sommadossi, J. P., & Richman, D. D. (2000). Drug Susceptibility in HIV Infection After Viral Rebound in Patients Receiving Indinavir-Containing Regimens. *JAMA*, 283(2), 229–234. <https://doi.org/10.1001/JAMA.283.2.229>
- Hikichi, Y., Duyne, R. Van, Pham, P., Groebner, J. L., Wiegand, A., Mellors, J. W., Kearney, M. F., & Freed, E. O. (2021). Mechanistic Analysis of the Broad Antiretroviral Resistance Conferred by HIV-1 Envelope Glycoprotein Mutations. *MBio*. <https://doi.org/10.1128/mBio>
- HIV.gov. (2021). *A Timeline of HIV and AIDS*. <https://www.hiv.gov/hiv-basics/overview/history/hiv-and-aids-timeline>

- Hocqueloux, L., Raffi, F., Prazuck, T., Bernard, L., Sunder, S., Esnault, J.-L., Rey, D., le Moal, G., Roncato-Saberan, M., André, M., Billaud, E., Valéry, A., Avettand-Fènoël, V., Parienti, J.-J., Allavena, C., & Study, M. (2019). Clinical Infectious Diseases Dolutegravir Monotherapy Versus Dolutegravir/Abacavir/ Lamivudine for Virologically Suppressed People Living With Chronic Human Immunodeficiency Virus Infection: The Randomized Noninferiority MONotherapy of TiviCAY Trial. *Clinical Infectious Diseases*®, 69(9), 1498–1505. <https://doi.org/10.1093/cid/ciy1132>
- Kearney, M. F., Spindler, J., Shao, W., Yu, S., Anderson, E. M., O’Shea, A., Rehm, C., Poethke, C., Kovacs, N., Mellors, J. W., Coffin, J. M., & Maldarelli, F. (2014). Lack of Detectable HIV-1 Molecular Evolution during Suppressive Antiretroviral Therapy. *PLoS Pathogens*, 10(3). <https://doi.org/10.1371/journal.ppat.1004010>
- Kleinpeter, A. B., & Freed, E. O. (2020). HIV-1 maturation: Lessons learned from inhibitors. In *Viruses* (Vol. 12, Issue 9). MDPI AG. <https://doi.org/10.3390/v12090940>
- la Rosa, A. M., Harrison, L. J., Taiwo, B., Wallis, C. L., Zheng, L., Kim, P., Kumarasamy, N., Hosseinipour, M. C., Jarocki, B., Mellors, J. W., & Collier, A. C. (2016). Raltegravir in second-line antiretroviral therapy in resource-limited settings (SELECT): a randomised, phase 3, non-inferiority study. *The Lancet. HIV*, 3(6), e247–e258. [https://doi.org/10.1016/S2352-3018\(16\)30011-X](https://doi.org/10.1016/S2352-3018(16)30011-X)
- Lee, G. Q., Reddy, K., Einkauf, K. B., Gounder, K., Chevalier, J. M., Dong, K. L., Walker, B. D., Yu, X. G., Ndung’u, T., & Lichterfeld, M. (2019). HIV-1 DNA sequence diversity and evolution during acute subtype C infection. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-019-10659-2>

Los Alamos National Library. (2017). *Landmarks of the HIV genome*.

<https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html>

Malet, I., Subra, F., Charpentier, C., Collin, G., Descamps, D., Calvez, V., Marcelin, A. G., & Delelis, O. (2017). Mutations located outside the integrase gene can confer resistance to HIV-1 integrase strand transfer inhibitors. *MBio*, *8*(5).

<https://doi.org/10.1128/MBIO.00922-17/ASSET/AAECBCCA-7EAF-4566-AC85-49E1B03887ED/ASSETS/GRAPHIC/MBO0051735020005.JPEG>

Manasa, J., Varghese, V., Kosakovsky Pond, S. L., Rhee, S.-Y., Tzou, P. L., Fessel, W. J., Jang, K. S., White, E., Rögnvaldsson, T., Katzenstein, D. A., & Shafer, R. W. (2017). Evolution of gag and gp41 in Patients Receiving Ritonavir-Boosted Protease Inhibitors. *Scientific Reports*. <https://doi.org/10.1038/s41598-017-11893-8>

McManus, W. R., Bale, M. J., Spindler, J., Wiegand, A., Musick, A., Patro, S. C., Sobolewski, M. D., Musick, V. K., Anderson, E. M., Cyktor, J. C., Halvas, E. K., Shao, W., Wells, D., Wu, X., Keele, B. F., Milush, J. M., Hoh, R., Mellors, J. W., Hughes, S. H., ... Kearney, M. F. (2019). HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. *Journal of Clinical Investigation*, *129*(11), 4629–4642.

<https://doi.org/10.1172/JCI126714>

Palmer, S., Kearney, M., Maldarelli, F., Halvas, E. K., Bixby, C. J., Bazmi, H., Rock, D., Falloon, J., Davey, R. T., Dewar, R. L., Metcalf, J. A., Hammer, S., Mellors, J. W., & Coffin, J. M. (2005). Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis.

*Journal of Clinical Microbiology*, *43*(1), 406–413. <https://doi.org/10.1128/JCM.43.1.406-413.2005>

- Palmer, S., Maldarelli, F., Wiegand, A., Bernstein, B., Hanna, G. J., Brun, S. C., Kempf, D. J., Mellors, J. W., Coffin, J. M., & King, M. S. (2008). *Low-level viremia persists for at least 7 years in patients on suppressive antiretroviral therapy*.  
[www.pnas.org/cgi/doi/10.1073/pnas.0800050105](http://www.pnas.org/cgi/doi/10.1073/pnas.0800050105)
- Saladini, F., Giannini, A., Boccuto, A., Tiezzi, D., Vicenti, I., & Zazzi, M. (2017). The HIV-1 integrase E157Q polymorphism per se does not alter susceptibility to raltegravir and dolutegravir in vitro. *AIDS*, *31*(16), 2307–2309.  
<https://doi.org/10.1097/QAD.0000000000001616>
- Salazar-Gonzalez, J. F., Bailes, E., Pham, K. T., Salazar, M. G., Guffey, M. B., Keele, B. F., Derdeyn, C. A., Farmer, P., Hunter, E., Allen, S., Manigart, O., Mulenga, J., Anderson, J. A., Swanstrom, R., Haynes, B. F., Athreya, G. S., Korber, B. T. M., Sharp, P. M., Shaw, G. M., & Hahn, B. H. (2008). Deciphering Human Immunodeficiency Virus Type 1 Transmission and Early Envelope Diversification by Single-Genome Amplification and Sequencing. *Journal of Virology*, *82*(8), 3952–3970. <https://doi.org/10.1128/JVI.02660-07>
- Sanders-Buell, E., Salminen, M. O., & Mccutchan, F. E. (1995). Sequencing Primers for HIV-1. *Los Alamos National Laboratory* , 1521–1527.
- Spach, D. (2022). *Evaluation and Management of Virologic Failure - Antiretroviral Therapy*. National HIV Curriculum . <https://www.hiv.uw.edu/go/antiretroviral-therapy/evaluation-management-virologic-failure/core-concept/all>
- Swanson, P., Devare, S. G., & Hackett, J. (2004). Molecular Characterization of 39 HIV-1 Isolates Representing Group M (Subtypes A-G) and Group O: Sequence Analysis of gag p24, pol Integrase, and env gp41. *https://Home.Liebertpub.Com/Aid*, *19*(7), 625–629.  
<https://doi.org/10.1089/088922203322231003>

- Van Duyne, R., Kuo, L. S., Pham, P., Fujii, K., & Freed, E. O. (2019). Mutations in the HIV-1 envelope glycoprotein can broadly rescue blocks at multiple steps in the virus replication cycle. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(18), 9040–9049. <https://doi.org/10.1073/pnas.1820333116>
- Van Zyl, G. U., Katusiime, M. G., Wiegand, A., McManus, W. R., Bale, M. J., Halvas, E. K., Luke, B., Boltz, V. F., Spindler, J., Laughton, B., Engelbrecht, S., Coffin, J. M., Cotton, M. F., Shao, W., Mellors, J. W., & Kearney, M. F. (2017). No evidence of HIV replication in children on antiretroviral therapy. *Journal of Clinical Investigation*, *127*(10), 3827–3834. <https://doi.org/10.1172/JCI94582>
- WHO. (2020). Global HIV & AIDS statistics. Fact sheet. In *Unaids*.
- Wiegand, A., Spindler, J., Hong, F. F., Shaoc, W., Cyktor, J. C., Cillo, A. R., Halvas, E. K., Coffin, J. M., Mellors, J. W., & Kearney, M. F. (2017). Single-cell analysis of HIV-1 transcriptional activity reveals expression of proviruses in expanded clones during ART. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(18), E3659–E3668. <https://doi.org/10.1073/pnas.1617961114>
- Wong, J. K., Hezareh, M., Günthard, H. F., Havlir, D. v., Ignacio, C. C., Spina, C. A., & Richman, D. D. (1997). Recovery of Replication-Competent HIV Despite Prolonged Suppression of Plasma Viremia. *Science*, *278*(5341), 1291–1295. <https://doi.org/10.1126/SCIENCE.278.5341.1291>