HIV-1 mutations in HIV-1 Gag, protease, RT p66 and when they appear: Insights from an *in vitro* BSL2 assay on mutation rates and types.

Joshua Yi Yeo[a], Ping Yap[a], Ghin-Ray Goh[a], Darius Wen-Shuo Koh[a], & Samuel Ken-En Gan[a,b,c,#]

[a]*Antibody & Product Development Lab, BII, Agency for Science, Technology and Research (A*STAR), Singapore 138672.*

[b]*Experimental Drug Development Centre, Agency for Science, Technology and Research (A*STAR), Singapore 138670*

[c]*p53 Laboratory, Agency for Science, Technology and Research (A*STAR), Singapore 138648.*

#Corresponding author: samgan@apdskeg.com; samuelg@bii.a-star.edu.sg

Antibody & Product Development Lab, BII, A*STAR

60 Biopolis Street, #B2 Genome

Singapore 138672

Tel: +65 64070584

Keywords: HIV, mutation rate, Gag, protease, reverse transcriptase, drug resistance

Abstract

The emergence of drug resistant mutants of HIV-1 is largely credited to its error prone HIV-1 RT, making it important to studying its fidelity and bias. Previous HIV-1 RT studies rely on reporter genes (LacZ) as its template, leaving a gap of understanding to how drug resistance mutations arise on HIV genes. This research aims to characterize HIV-1 RT mutations on the Gag, protease, and RT p66 subunit in an *in-vitro* selection pressure free assay for studying the emergence of drug resistance mutations. Our results showed that known drug resistance mutations occurred early with a general avoidance of key sites and detrimental mutations in the backdrop of mutations rates of 4.71×10^{-5}, 6.03×10^{-5}, and 7.09×10^{-5} mutations/bp for Gag, protease and RT p66 respectively, with significant 'A to G' hypermutations likely influenced by cellular adenosine deaminases. Silently mutating RT p66 to study the effect of sequence on mutation rate and type, we observed an increase in both rate (1.89×10^{-4} mutations/bp) and 'A to G' hypermutations. Our study demonstrates the importance of studying mutation emergence in HIV genes to understand how fast drug resistance can emerge, and in this, provide potential horizontal understanding to other viruses to understand how new viral disease and drug resistance can emerge.

Introduction

Emerging viral infections are likely to be RNA viruses which are shown to have a higher likelihood of undergoing genetic changes, leading to species jump[1] and efficient spread among humans[2]. Among the RNA viruses, the simian immunodeficiency virus (SIV) lineages are believed to have ancestral links to the human immunodeficiency virus, HIV-1 and HIV-2[3,4].

HIV-1 Gag and protease play vital roles in viral assembly and maturation for the production of infectious virions[5]. HIV-1 protease cleaves HIV-1 Gag and pol polyproteins into functional subunits[6]. To prevent maturation, protease inhibitors (PI) bind competitively against Gag to the active site of HIV-1 protease, thereby inhibiting viral maturation[7,8]. To overcome PIs, drug resistance mutations on viral protease reduce affinity to PIs[8,9], and the accumulation of such resistance mutations[10-12] induced by HIV-1 RT[13,14] can render PIs ineffective and confer cross-resistance, limiting drug selection[8]. In addition, mutations have been reported on HIV-1 Gag to compensate for reduced viral fitness[8,9,14,15], with paired-mutations that work synergistically in both HIV-1 Gag and protease also reported[16-20].

The key to the generation of mutations in HIV genes is by the error prone HIV-1 reverse transcriptase (RT), an asymmetric heterodimer consisting of the p66 and p51 subunits[21]. The p66 subunit performs the key enzymatic functions, catalysing DNA polymerisation and cleaving the RNA of the RNA/DNA duplex[22,23], while the p51 subunit plays a supportive role to p66[24].

Since most of the previous studies on HIV mutation rate were based on reporter genes (LacZ) and not HIV genes, there is a gap in the understanding of HIV RT mutation rates on HIV genes and when drug resistance mutations are generated[25]. To fill this gap, we developed an *in-vitro* cell-based assay without

translational, immune and drug selection pressures to characterize the basal HIV-1 RT mutations and biases on HIV-1 Gag, protease, RT p66 genes, and a sequence variant of RT p66 that had all codons silently mutated to investigate the effect of codon usage on mutation rate. This *in vitro* assay allowed us to establish the native error rate of HIV-1 RT to establish the rate, position and bias in HIV-1 Gag, protease, RT p66 in the emergence of drug resistance.

Results

Characterization of HIV-1 Gag mutant variants

In our study, we classified 'A to G' mutations as hypermutations when multiple mutations were found on the same sequence variant given that mutations are expected to be consistently spread out.

We generated a total of 801 HIV-1 Gag sequences and found a mutation rate of 3.36×10^{-5} mutations/bp when excluding 'A to G' hypermutations and a slightly higher rate of 4.71×10^{-5} mutations/bp when including them, assuming the experiment setup successfully mimicked one replication cycle (see Table 1).

There were more transitions (n = 24 excluding 'A to G' hypermutations, n = 40 when including 'A to G' hypermutations) as compared to transversions (n = 12) as shown in Figure 1A. For nucleotide substitutions, 'A to G' substitutions were found to be the highest (n = 20, 37.7%) when including 'A to G' hypermutations. When excluding 'A to G' hypermutations, 'T to C' substitutions were the highest (n = 9, 25.0%), followed by 'G to A' and 'G to T' (n = 6, 16.7%), 'C to A' and 'C to T' (n = 5, 13.9%), 'A to G' (n = 4, 11.1%), and 'T to G' (n = 1, 2.8%), with several substitutions not observed (see Figure 1B and 1C).

Analysis at the amino acid level showed missense mutations (n = 30, 55.6%) to occur at approximately 2.0 times the frequency of silent mutations (n = 15, 27.8%,

see Figure 2) while nonsense mutations (n = 5, 9.3%) and frameshift mutations (n = 4, 7.5%) had lower occurrences.

Characterization of HIV-1 protease mutant variants

For HIV-1 protease, we generated a total of 640 sequences, finding a mutation rate of 6.03 x 10^{-5} mutations/bp (see Table 1), noting that no 'A to G' hypermutations were found for protease.

As with Gag, there were more transition mutations (n = 6) than transversions (n = 2) as shown in Figure 1A. 'G to A' substitutions were found to be the highest (n = 3, 37.5%), followed by 'A to G' (n = 2, 25%), 'A to C', 'T to C' and 'C to A' (n = 1, 12.5%), with several substitutions not observed (see Figure 1B and 1C).

At the amino acid level, missense mutations (n = 4, 36.4%) occurred at approximately 1.3 times the frequency of silent mutations (n = 3, 27.3%) and frameshift mutations (n = 3, 27.3%), whereas nonsense mutations (n = 1, 9.1%) were of lower occurrences (see Figure 2).

Characterization of HIV-1 RT p66 mutant variants

From 571 HIV-1 RT p66 subunit sequences, we calculated a mutation rate of 4.27 x 10^{-5} mutations/bp when excluding 'A to G' hypermutations and 7.09 x 10^{-5} mutations/bp when including them (see Table 1).

As with Gag and protease, there were more transition mutations (n = 18 excluding 'A to G' hypermutations, 45 including 'A to G' hypermutations) as compared to transversions (n = 11) as shown in Figure 1A. 'A to G' substitutions were found to be the highest (n = 29, 51.8%) only when including 'A to G' hypermutations. When 'A to G' hypermutations were excluded, 'T to C' and 'G to T' substitutions were instead the highest occurrences (n = 8, 27.6%), followed by 'G to

A' (n = 6, 20.7%), 'C to A' (n = 4, 12.5%), and 'A to G' and 'C to T' (n = 2, 6.9%). There were several substitutions not observed (see Figure 1B and 1C).

Analysis after translation found relative consistency with Gag and protease where missense mutations (n = 33, 49.3%) occurred at 1.7 times the frequency of silent mutations (n = 20, 29.9%) as shown in Figure 2. Nonsense mutations (n = 2, 3.0%) and frameshift mutations (n = 12, 17.9%) were of lower occurrences.

Characterization of HIV-1 Codon Mutated RT p66 mutant variants

A total of 700 codon mutated p66 gave a calculated mutation rate of 6.63 x 10^{-5} mutations/bp when excluding 'A to G' hypermutations and 1.89 x 10^{-4} mutations/bp when including them (see Table 1).

Consistent with the earlier mentioned genes, there were more transition mutations (n = 38 excluding 'A to G' hypermutations, 177 including 'A to G' hypermutations) as compared to transversion (n = 29) as shown in Figure 1A. As with the wild-type p66, 'A to G' substitutions were found to be the highest (n = 139, 70.9%) only when including 'A to G' hypermutations. Excluding 'A to G' hypermutations, 'G to T' substitutions were found to be the highest (n = 17, 25.4%), followed by 'G to A' (n = 14, 20.9%), 'T to C' (n = 13, 19.4%), 'A to G' (n = 7, 10.4%), 'C to A' (n = 8, 11.9%), 'C to T' (n = 4, 6.0%), 'A to T' (n = 2, 3.0%), and 'T to A' and 'T to G' substitutions (n = 1, 1.5%), with the absence of the following substitutions: 'A to C', 'C to G' and 'G to C' (see Figure 1B and 1C).

At the amino acid level, relative consistency with the other HIV-1 genes was also observed with missense mutations (n = 109, 56.2%) occurring at approximately 1.6 times the frequency of silent mutations (n = 70, 36.1%, see Figure 2). Nonsense mutations (n = 7, 3.6%) and frameshift mutations (n = 8, 4.1%) were at lower occurrences.

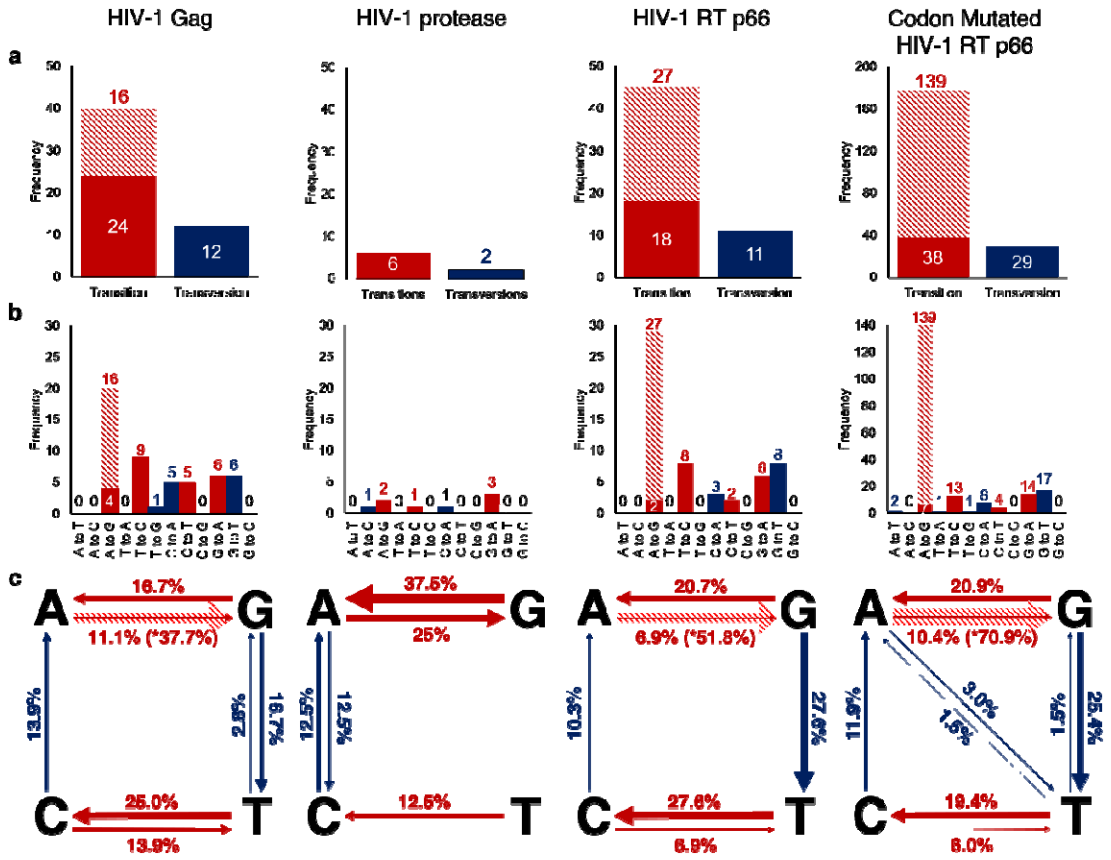

Fig. 1: Generated HIV-1 Gag, protease, RT p66 and Codon Mutated RT Mutated p66 nucleotide substitution mutations using HIV-1 RT.

a Bar chart of the transition and transversion mutations in HIV-1 Gag (n = 52), protease (n = 8), RT p66 (n = 56) and codon mutated p66 (n = 206). Transition and transversion mutations are shown in red and blue respectively. 'A to G' Hypermutations are shown separately in a diagonal striped pattern. **b** Bar chart showing substitution mutations in HIV-1 Gag, protease, RT p66 and codon mutated p66 sequences. **c** Schematic diagram of nucleotide substitutions occurring on the respective HIV-1 genes (expressed as a percentage).

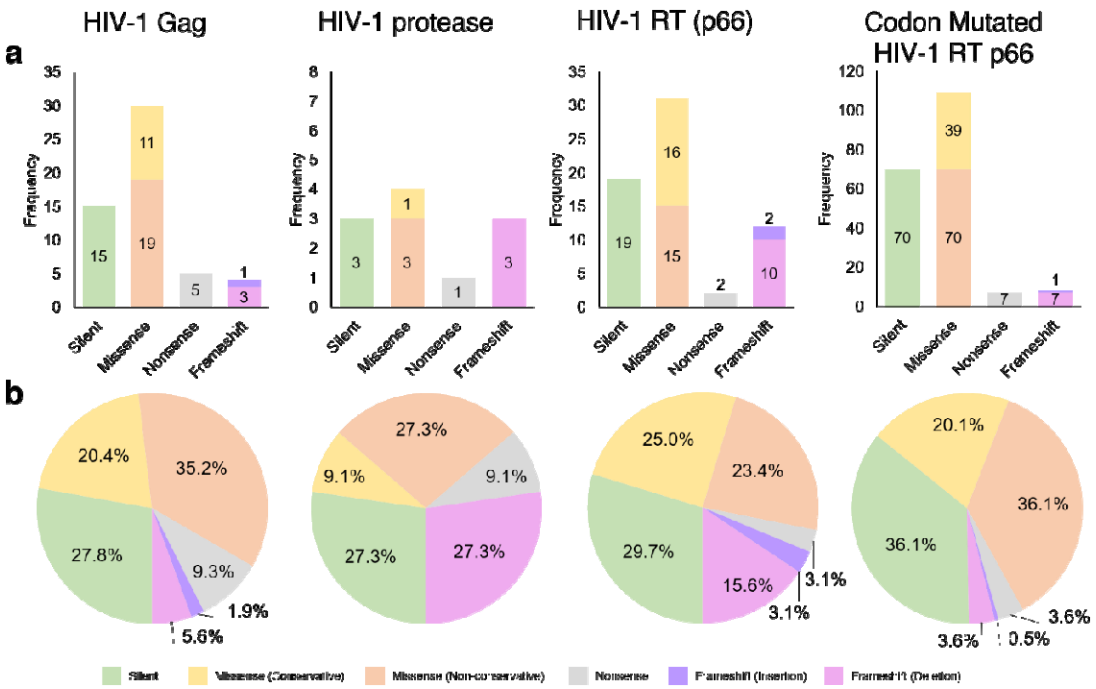

Fig. 2: Generated amino acid mutations using HIV-1 RT.

a Bar chart of amino acid mutations observed in HIV-1 Gag (n = 54), protease (n = 11), RT p66 (n = 64) and codon mutated RT p66 (n = 194). **b** Pie chart of the percentage proportions of the type of translation effects resulting from the mutations. For the codon mutated RT p66, two variants (20 and 80 a,b) containing deletions without frameshifts are not shown. Silent, missense (conservative), missense (non-conservative), nonsense, frameshift (insertion) and frameshift (deletion) mutations are shown in green, yellow, orange, grey, purple and pink, respectively.

Table 1: Calculated error rates of HIV-1 RT on the respective HIV-1 genes

HIV-1 Gene	No. of clones	Nucleotide Length	Total no. of bases	Excluding 'A to-G' Hypermutations		Including 'A to G' Hypermutations	
				No. of mutations	Error rate (mutations/bp)	No. of mutations	Error rate (mutations/bp)
Gag	801	1485	1,189,485	40	3.36×10^{-5}	56	4.71×10^{-5}
protease	640	285	182,400	11	6.03×10^{-5}	-	-
RT p66	571	1680	959,280	41	4.27×10^{-5}	68	7.09×10^{-5}
Codon mutated RT p66	700	1617	1,131,900	75	6.63×10^{-5}	214	1.89×10^{-4}

Error rates were calculated as the ratio of the total number of mutations and the total number of nucleotide bases, comparable to the reported range of 1.8×10^{-5} – 6.67×10^{-4}, previously reviewed[25,26].

Mutation Rates of the Domains of HIV-1 Gag, protease, RT and RT p66

The HIV-1 genes were sorted into their respective domains and the mutation rates calculated (see Figure 3). Within Gag, the overall mutation rates throughout the domains were within a fairly narrow range of 2.29 to 4.46 x 10^{-5} mutations/bp when excluding 'A to G' hypermutations, and increasing to 8.16 x 10^{-5} when including them. 'A to G' hypermutations were found in the capsid (CA), nucleocapsid (NC), and p6 subunits, but not the matrix (MA), p2 and p1 domains. Mutations were not observed in the MA/CA ^{129}SQNY/PIV^{135}, CA/p2 ^{360}ARVL/AEA^{366}, p2/NC ^{374}ATIM/IQK^{380}, P1/p6 ^{445}PGNF/LQS^{451} cleavage sites and the ^{277}YSPTSIL^{283} capsid linker. However, A431D and F433L were observed in the NC/p1 ^{429}RQAN/FLG^{435} cleavage site.

For HIV-1 protease, there were no 'A to G' hypermutations throughout the gene and no mutations observed in the active site ^{25}DTG^{27}. For RT p66, there was a range of 0 – 9.12 x 10^{-5} mutations/bp excluding hypermutations, which increased to 1.35 x 10^{-4} mutations/bp when including them. Despite the 2nd finger domain being of similar length as the 1st palm segment, only 'A to G' hypermutations were observed without other mutations. In the RT p51 and RT RNase H Domain (p51-RNH) cleavage site AETF^{440}↓Y^{441}VD and catalytic triad D110, D185 and D186, there were also no mutations observed. In the codon mutated p66, the mutation rate was in a range of 2.98 x 10^{-5} – 1.06 x 10^{-4} mutations/bp when excluding hypermutations which increased the rate to 3.38 x 10^{-4} mutations/bp. Contrary to the wild-type p66, there were mutations observed in the 2nd finger domain, suggesting that codon usage of the wild-type 2nd finger domain had some barrier to mutations. Similarly, no mutations were observed in the p51-RNH cleavage site AETF^{440} ↓ Y^{441}VD despite

the different codon usage. Interestingly, ‘A to G’ hypermutations were not observed in the 1st palm region of the codon mutated p66.

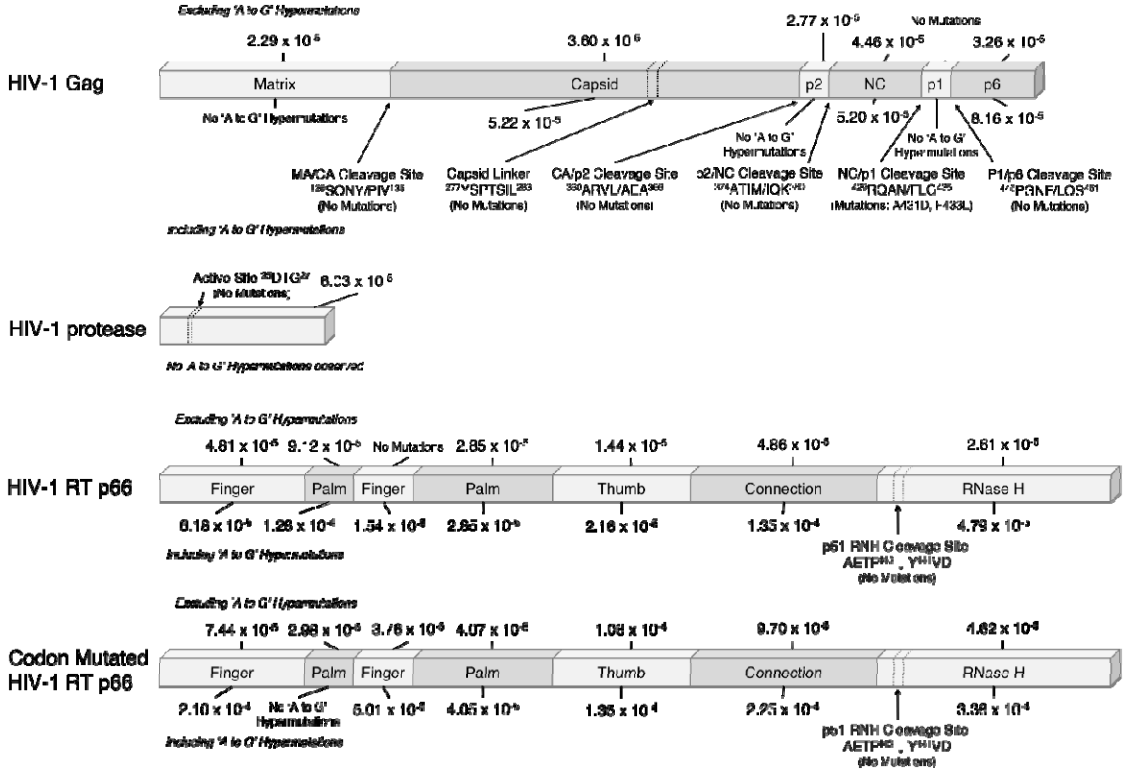

Fig. 3: Mutation rates (mutations/bp) calculated in the domains of HIV-1 Gag, protease, RT p66 and Codon Mutated RT p66.

Mutation rates were calculated as the ratio of the total number of mutations and the total number of nucleotide bases, and counted separately excluding and including A to G hypermutations. Truncations that span across multiple domains were excluded in the calculations. No mutations were generated in the MA/CA ^{129}SQNY/PIV^{135}, CA/p2 ^{360}ARVL/AEA^{366}, p2/NC ^{374}ATIM/IQK^{380}, P1/p6 site ^{445}PGNF/LQS^{451} cleavage sties and ^{277}YSPTSIL^{283} for HIV-1 Gag, nor in the catalyric triad ^{25}DTG^{27} in HIV-1 protease, nor in the p51-RNH cleavage site AETF^{440} ↓ Y^{441}VD and catalytic triad D110, D185 and D186 in both HIV-1 RT p66 and codon mutated HIV-1 RT p66.

A431D and F433L were observed in the NC/p1 cleavage site ^{429}RQAN/FLG^{435} of HIV-1 Gag.

Discussion

We set out to study the mutation rate of HIV-1 RT on HIV-1 Gag, protease and RT p66 subunit in a single replication cycle using our *in vitro* BSL2 assay devoid of selection pressures to establish the native mutation rate of HIV-RT on HIV genes. Given the absence of translational, immune, fitness, and drug selection pressures in our system, the observed biases are presumed to be intrinsic to HIV-RT and due to host cell factors. Although mutations can be contributed from plasmid hosts (DH5α and/or HEK293F), Q5 polymerase and sequencing artefacts, these were unlikely given that the observed phenotypic mutations/bp/replication for *E. coli* and mammalian cells were estimated to be 5.4×10^{-10} and 5.0×10^{-11} respectively[27], and that Q5 polymerase is among the most accurate commercial high-fidelity polymerase[28]. Admittedly, there may be a role for primer selection bias when priming the first few bases of the genes, and as such, these bases were excluded from analysis and calculation of mutation rates.

Although numerous *in vitro* studies have attempted to establish the error rate of HIV-1 RT (as previously reviewed[25,26]), merely two utilised HIV-1 genes as their template[29,30]. In our assay, we found the mutation rates of HIV-1 Gag, protease, RT p66 and codon mutated p66 to be 4.71×10^{-5}, 6.03×10^{-5}, 7.09×10^{-5} and 1.89×10^{-4} mutations/bp (inclusive of hypermutations), respectively, within the range of 1.8×10^{-5} – 6.67×10^{-4} mutations/bp previously reported[25,26]. Across the HIV-1 genes, we observed a predominance of transition mutations that is consistent with the characterization across most phyla, previously hypothesized to conserve protein functions[31-33]. Comparing the types of amino acid introduced, we observed a general trend with missense mutations having the highest frequency, followed by silent mutations, frameshifts and nonsense. This is notable because even in the absence

of selection pressures to maintain protein translatability, frameshift and nonsense mutations (with more detrimental effects) were occurring at a lower frequencies, possibly constrained by in-built genetic code probabilities[34].

When changing the HIV-1 gene templates, a change in mutation rates, type of mutations and mutational biases was also observed. In general, there were no A – T or C – G mutations in the HIV genes studied, which were only detected in the codon mutated p66 (see Figure 1), suggesting influence of the sequence or codon usage. Comparing the HIV-1 codon mutated RT p66 to wild-type HIV-1 RT p66, the mutation rate (1.89×10^{-4} mutations/bp) increased by 2.7 folds as compared to 7.09×10^{-5}. Interestingly, the RNase H domain of the codon mutated HIV-1 RT p66 was also found to have a 7.1 folds increase (3.38×10^{-4} mutations/bp), as compared to the wild-type (4.79×10^{-5} mutations/bp) suggesting RNA secondary structural protection/susceptibility to host adenosine deaminases[35-37].

Of the genes, the 'A to G' hypermutations were observed in HIV-1 Gag, RT p66 and codon mutated RT p66, but not in protease, possibly due to its short nucleotide length. In this, there is an interesting possibility that smaller genes avoided the effects of A to G mutational effects that are likely caused by host adenosine deaminases such as double-stranded RNA-specific adenosine deaminase (ADAR) that deaminates 'A' residues to inosine (I) residues (read as 'G' by ribosomes due to structural homology), leading to 'C' misincorporations in the plus strand DNA after reverse transcription, and thus resulting in 'G' residues being incorporated during second (minus) stand synthesis[38,39]. As such, ADAR has been shown to display both pro-viral and anti-viral activity[40-43].. Given the lack of apparent ADAR effects on protease, protease is likely the better target than HIV-1 RT or Gag,

which may explain the lower occurrences of PI resistance in clinical studies[44], alongside better potency than reverse transcriptase inhibitors (RTIs)[45].

With ‘A to G’ hypermutations, there would be a general accumulation towards G residues mutations for HIV-1 Gag, RT p66 and codon mutated RT p66, with the exception of protease. Such ‘G’ biasness would result in increased glycine, arginine, valine and alanine occurrence by the probability of codons[34]. On the contrary, this probability will result in reduced phenylalanine, isoleucine, tyrosine, histidine and asparagine presence in these genes. For HIV-1 protease, the mutational bias is towards ‘A’ and ‘C’, which is also the case when ‘A to G’ hypermutations are not considered for Gag and both p66 genes. The accumulation of ‘A’ would lead to a bias towards threonine and lysine, isoleucine, asparagine, arginine and stop codons, while reducing phenylalanine, tryptophan and cysteine occurrences. Similarly, the accumulation of ‘C’ would lead to proline, serine, leucine, threonine, alanine and arginine while reducing methionine, lysine, glutamic acid, tryptophan, and more importantly, stop codons. Notably, the mutational biases are trends, thus requiring the consideration of time to better characterise mutations, especially when considering how the influence of ADAR, together with such innate RT biases, provide the diversity observed for HIV.

Upon mapping the mutations observed, we did not find mutations in majority of the active sites of the enzymes (protease and RT p66) and cleavage sites of Gag, with the exception of A431D and F433L in NC/p1 cleavage site ^{429}RQAN/FLG^{435} of Gag, suggesting that cleavage compensation mutations coud have emerged early in the infection process even in the absence of drug seletion. Despite being of a single replication cycle, our assay also detected known clinical drug resistance mutations (see Supplementary Table 5 for the classified ones on Gag, protease, and p66). For

instance, in HIV-1 Gag, the observed E17K mutation (in the matrix) was previously reported as a rare transient mutation implicated in cytotoxic T lymphocyte (CTL) immune evasion resistance[46,47]. In HIV-1 protease, the observed K70T, previously reported to be a minority mutation associated with resistance to PIs[48] was also found in our assay. In HIV-1 RT p66, a known polymorphic mutation K103R (in the palm domain) was observed, which when in combination with V179D (not found in our assay), reduced susceptibility to non-nucleoside reverse transcriptase inhibitors (NNRTIs): nevirapine (NVP) and efavirenz (EFV) by about 15 folds[49]. Interestingly, the observed K103R originated from an 'A to G' mutation, specifically from "AAA" to "AGA", suggesting that drug resistance mutations can be induced by host cell ADAR.

Few mutations were also found in crucial functional sites within the genes. In HIV-1 Gag, we found P222L in the cyclophilin A (CyPA) binding site (along with G221) that might potentially affect binding of the capsid to CyPA as previously reported for P222A[50]. Given that the probability of P (CCT) to L (CTT) is 0.33 in having a T substitution in the 2nd codon position, and that P (CCT) to A (GCT) is also 0.33 in having the first C mutated to a G, there should be equal probabilities. However, when considering that our Gag mutation analysis did not shown a transition mutation link between C to G (shown in Figure 1), it is likely that the clinical mutation is a rare-occurrence (under additional selection pressures like drug) or a two or more mutation steps. Mutations A431D in the nucleocapsid and P453T in the p6 domain were also observed, with reported clinical drug resistance counterparts as A431V and a L449F/P453T pair[51]. In HIV-1 RT p66, the substitution F61S was observed, in which the F61 residue was previously reported to play a vital role in strand synthesis, with reported corresponding counterparts F61Y/L/W altering activity[52]. Similarly, we observed P95L, in which P95 was reported to be at the

dimerization interface for the formation of the bottom of NNRTI pocket[53], and previously proposed to be a target amino acid in NNRTI design together with N137 and P140[54]. Interestingly, although P95 was reported to be in a highly conserved location[53], it was observed in our assay.

The occurrence of previously reported *in vitro* and clinical mutations in our assay suggest that these mutations occurred without immune surveillance and drug selection pressure (such as ART), occurring by probability within HIV gene specific biases of RT (Figure 1). The observation of a general avoidance of inducing mutations at key regions/sites even at the nucleotide stage, seems to suggest in-built barriers to detrimental mutations. Most relevant to the clinic is that these mutations occurred within our mimic of a single cycle of replication. Considering that HIV-1 generates approximately 10^9 virions per day in an infected individual[55] which is magnitudes higher than our hundreds of sequence variants per gene, clearly drug resistance isolate can occur even within the first day of infection.

In the absence of selection pressures, mutations that are selected against can be better detected, allowing for a more comprehensive analysis of the HIV-1 RT bias. Using current methods, including the high-throughput ones like next-generation sequencing, low frequency mutations can be missed to only emerge during intense selection. Given that many HIV proteins can function in intense drug/immune selection environments even with significant reduced activity[56,57], an all-out campaign against HIV could involve inhibitors against low-frequency occurring drug resistance mutations. This pushes the HIV infection towards clearance by Muller's ratchet[58-60] or by increasing viral mutation rate beyond pass a critical threshold towards lethal mutagenesis, leading to the loss in genetic information and viability, error catastrophe[61,62]. Since HIV-1 replicates at the edge of this critical threshold[63], a

slight increase in the mutation rate may be sufficient to bring about error catastrophe, which have been successfully induced previously by mutagenic nucleoside analogues[64-66].

Coupling with *in silico* tools, such as those incorporating probability mutational change, it may be possible to identify mutational hotspots. As it was previously shown that protease and RT drug cross resistance have a structural basis governed by drug resistance mutations [67,68], the bias of restricting specific amino acid changes by the absence of A-T, C-G mutation occurrences can be exploited to constrain the emergence of drug resistance. However, such an approach will require an in-depth understanding of HIV-RT mutations that do not proceed to a protein level under the most basic biological selection pressures of protein function.

Through the understanding of the mutational biases and mutation rate of HIV-1 RT, it possible to calculate the mutational events leading to the speciation jump from an early SIV to the current HIV, making it important for surveillance for emerging viral threats[2]. Although species jump is unpredictable, RNA viruses are most likely to species jump after undergoing genetic changes[1].

In conclusion, we have established an assay to generate and characterise novel HIV-1 RT mutations on HIV-1 Gag, protease and RT p66 in a safe BSL2 manner, allowing insights to the mutational bias and mutation rate of HIV in the absence of biological selection pressures. In our analysis, we also characterized the likely effect of host factors e.g. ADAR which increased the genetic diversity beyond the already error-prone RT. In this genetic diversity, we showed that we were able to generate known clinical drug resistance mutations while mimicking a single replication cycle, showing that these mutations would be present very early in

infection. The assay can provide deeper insights relevant for drug and vaccine development and be applied for horizontal understanding to other viruses.

Methods

Transfection of HIV-1 Gag, PR and RT plasmids

HIV-1 Gag and PR genes were PCR amplified from plasmid p8.91[69]. HIV-1 RT p66 (GenBank: K03455.1) and a codon mutated version of RT p66 was gene synthesized (BioBasic). Codon mutated p66 was generated by reverse translation from the amino acid sequence that would give rise to silent mutations with differences at the nucleotide sequence level (sequence available in Supplementary Figure 4) to disrupt RNA secondary structures. The genes were cloned separately into pTT5 plasmid vector (YouBio), transformed into in-house competent DH5α *E. coli* cells[70], and transfected into HEK293F cells as previously performed[71-73].

RNA extraction and cDNA synthesis

Total RNA was extracted using TRIzol reagents according to manufacturer's instructions (Invitrogen) from transfected cells two days after transfection with the gene of interest. cDNA synthesis was performed using commercial HIV-1 RT subunits: (i) p51 (0.2475 µg) and p66 (0.2125 µg) from Sino Biological Inc (catalogue: 40244-V07E and 40244-V07E1, respectively), (ii) DNase-treated RNA, (iii) 5X RT buffer (25 mM Dithiothreitol, 375 mM KCl, 15 mM $MgCl_2$, 250 mM Tris-HCl [pH 8.3]), (iv) Oligo$(dT)_{18}$ (Thermo Scientific), (v) dNTPs, and (vi) RiboLock RNase inhibitor (ThermoFisher Scientific), in a single cycle of 25°C for 18 min, 37°C for 1 hr and 85°C for 5 min. RT negative controls were prepared without the addition of HIV-1 RT p51 and p66 subunits.

Amplification of cDNA and TOPO cloning

High-fidelity Q5 Polymerase PCR (New England Biolabs) was used to amplify cDNA templates using the following primers for the respective genes: HIV-1 Gag - forward (5 '– TAT TAG GAA TTC ATG GGT GCG AGA GCG– 3') and reverse (5' – CTG GTA AAG CTT CTA GTG GTG GTG GTG – 3'); for protease - forward (5' – GCG GCC GAA TTC ATG CCT CAA ATC AC – 3') and reverse (5' – TAT AAT AAG CTT CTA GTG GTG GTG GTG –3'); for RT p66 - forward (5' – ATG GCC TTG ACC TTT GCT TTA CTG – 3') and reverse (5'– CTT GTC GTC ATC GTC TTT GTA GTC – 3'); for codon mutated p66 - forward (5'– GCG GTG ATG GAT GGA CCA AAA GTA AA – 3') and reverse (5'– CTG CGC CTA ATG ATG ATG ATG AT – 3') as per manufacturer's recommendations. PCR products were analysed by gel electrophoresis using GelApp[74] and purified using Gel extraction and PCR purification kits as previously described[75]. Purified Q5 Polymerase PCR products were cloned using Zero Blunt TOPO PCR cloning kit (Invitrogen) as per manufacturer's protocol and transformed into in-house competent DH5α cells as previously described[70]. Transformed DH5α cells were plated and grown overnight at 37°C on LB agar plates supplemented with kanamycin (50 μg/ml). Transformants were screened using GoTaq PCR (Promega) with universal M13F (-20) forward and reverse primers prior to sequencing.

Sequence analysis

Sequence assembly and alignment of HIV-1 sequences were performed using YAQAAT Webserver (yaqaat.apdskeg.com). DNA2App[76] was used to translate nucleotide sequences to their corresponding amino acid sequences. Previously characterised HIV-1 sequences were retrieved from the Los Alamos sequence

database[77]. Mutations were identified through multiple sequence alignments with the transfected sequence. To rule out sequencing artefacts, sequence chromatograms were analysed, and ambiguous peaks and mutated sequences were sent for re-sequencing. Mutation rates were calculated as mutations/bp where the total number of mutations and the total number of nucleotide bases of the respective HIV-1 genes were compared.

Competing Interests

There are no conflicts of interest

Acknowledgments

We thank Wai-Heng Lua and Wei-Li Ling for useful comments and discussion.

Funding

This work was supported by the A*STAR Industry Alignment Fund (IAF) grant IAF111149 and by EDDC, A*STAR.

Author Contributions

JYY, PY, and GRG performed the *in vitro* experiments. JYY, PY, GRG, and DWSK analysed the results. JYY, PY, GRG, DWSK, and SKEG discussed and drafted the manuscript. SKEG conceived the idea and supervised the whole project. All authors read and approved the manuscript.

References

1 Woolhouse, M. E. J., Haydon, D. T. & Antia, R. Emerging pathogens: the epidemiology and evolution of species jumps. *Trends in Ecology & Evolution* **20**, 238-244, doi:10.1016/j.tree.2005.02.009 (2005).

2 Flanagan, M. L. *et al.* Anticipating the Species Jump: Surveillance for Emerging Viral Threats. *Zoonoses and Public Health* **59**, 155-163, doi:10.1111/j.1863-2378.2011.01439.x (2012).

3 Wertheim, J. O. & Worobey, M. Dating the Age of the SIV Lineages That Gave Rise to HIV-1 and HIV-2. *PLOS Computational Biology* **5**, e1000377, doi:10.1371/journal.pcbi.1000377 (2009).

4 Zhu, T. *et al.* An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391**, 594-597, doi:10.1038/35400 (1998).

5 Freed, E. O. HIV-1 assembly, release and maturation. *Nature Reviews* **13**, 484-496, doi:10.1038/nrmicro3490 (2015).

6 Pettit, S. C., Everitt, L. E., Choudhury, S., Dunn, B. M. & Kaplan, A. H. Initial Cleavage of the Human Immunodeficiency Virus Type 1 GagPol Precursor by Its Activated Protease Occurs by an Intramolecular Mechanism. *Journal of Virology* **78**, 8477-8485, doi:10.1128/jvi.78.16.8477-8485.2004 (2004).

7 Fun, A. *et al.* HIV-1 protease inhibitor mutations affect the development of HIV-1 resistance to the maturation inhibitor bevirimat. *Retrovirology* **8**, 70, doi:10.1186/1742-4690-8-70 (2011).

8 Fun, A., Wensing, A. M. J., Verheyen, J. & Nijhuis, M. Human Immunodeficiency Virus gag and protease: partners in resistance. *Retrovirology* **9**, 63, doi:10.1186/1742-4690-9-63 (2012).

9 Clavel, F. & Mammano, F. Role of Gag in HIV Resistance to Protease Inhibitors. *Viruses* **2**, 1411-1426, doi:10.3390/v2071411 (2010).

10 Wensing, A. M. *et al.* 2017 Update of the Drug Resistance Mutations in HIV-1. *Topics in Antiviral Medicine* **24** (2017).

11 Park, J. *et al.* Binding of clinical inhibitors to a model precursor of a rationally selected multidrug resistant HIV-1 protease is significantly weaker than that to the released mature enzyme. *Biochemistry* **55**, 2390-2400 (2016).

12 Yu, Y., Wang, J., Shao, Q., Shi, J. & Zhu, W. Effects of drug-resistant mutations on the dynamic properties of HIV-1 protease and inhibition by Amprenavir and Darunavir. *Sci Rep* **5**, 10517 (2015).

13 Su, C. T.-T., Kwoh, C.-K., Verma, C. S. & Gan, S. K.-E. Modeling the full length HIV-1 Gag polyprotein reveals the role of its p6 subunit in viral maturation and the effect of non-cleavage site mutations in protease drug resistance. *Journal of Biomolecular Structure and Dynamics*, 1-12, doi:10.1080/07391102.2017.1417160 (2017).

14 Clutter, D. S., Jordan, M. R., Bertagnolio, S. & Shafer, R. W. HIV-1 drug resistance and resistance testing. *Infection, Genetics and Evolution* **46**, 292-307, doi:10.1016/j.meegid.2016.08.031 (2016).

15 Dam, E. *et al.* Gag Mutations Strongly Contribute to HIV-1 Resistance to Protease Inhibitors in Highly Drug-Experienced Patients besides Compensating for Fitness Loss. *PLoS Pathogens* **5**, e1000345, doi:10.1371/journal.ppat.1000345 (2009).

16 Parry, C. M. *et al.* Gag Determinants of Fitness and Drug Susceptibility in Protease Inhibitor-Resistant Human Immunodeficiency Virus Type 1. *Journal of Virology* **83**, 9094-9101 (2009).

17 Dam, E. *et al.* Gag mutations strongly contribute to HIV-1 resistance to Protease Inhibitors in highly drug-experienced patients besides compensating for fitness loss. *Plos Pathogen* **5**, e1000345 (2009).

18 Tamiya, S., Mardy, S., Kavlick, M. F., Yoshimura, K. & Mitsuya, H. Amino acid insertions near Gag cleavage sites restore the otherwise compromised replication of human immunodeficiency virus type 1 variants resistant to Protease Inhibitors. *J Virol* **78**, 12030-12040 (2004).

19 Maguire, M. F. *et al.* Changes in Human Immunodeficiency Virus Type 1 Gag at Positions L449 and P453 Are Linked to I50V Protease Mutants In Vivo and Cause Reduction of Sensitivity to Amprenavir and Improved Viral Fitness In Vitro. *Journal of Virology* **76**, 7398-7406 (2002).

20 Gatanaga, H. *et al.* Amino Acid Substitutions in Gag Protein at Non-cleavage sites Are Indispensable for the Development of a high Multitude of HIV-1 Resistance against Protease Inhibitors. *The journal of Biological Chemistry* **277**, 5952-5961 (2002).

21 Lightfoote, M. M. *et al.* Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus. *Journal of virology* **60**, 771-775 (1986).

22 Hostomsky, Z., Hostomska, Z., Fu, T. B. & Taylor, J. Reverse transcriptase of human immunodeficiency virus type 1: functionality of subunits of the heterodimer in DNA synthesis. *J Virol* **66**, 3179-3182 (1992).

23 Sarafianos, S. G. *et al.* Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol* **385**, 693-713, doi:10.1016/j.jmb.2008.10.071 (2009).

24 Le Grice, S. F., Naas, T., Wohlgensinger, B. & Schatz, O. Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1 reverse transcriptase. *EMBO J* **10**, 3905-3911 (1991).

25 Yeo, J. Y., Goh, G. R., Su, C. T.-T. & Gan, S. K.-E. The Determination of HIV-1 RT Mutation Rate, Its Possible Allosteric Effects, and Its Implications on Drug Resistance. *Preprints*, 2020020099, doi:doi: 10.20944/preprints202002.0099.v1 (2020).

26 Menéndez-Arias, L. Mutation Rates and Intrinsic Fidelity of Retroviral Reverse Transcriptases. *Viruses* **1**, 1137-1165 (2009).

27 Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics* **148**, 1667-1686 (1998).

28 Potapov, V. & Ong, J. L. Examining sources of error in PCR by single-molecule sequencing. *PloS One* **12**, e0169774 (2017).

29 Geller, R. *et al.* The external domains of the HIV-1 envelope are a mutational cold spot. *Nature Communications* **6**, 8571, doi:10.1038/ncomms9571 (2015).

30 Ji, J. & Loeb, L. A. Fidelity of HIV-1 Reverse Transcriptase Copying a Hypervariable Region of the HIV-1 env Gene. *Virology* **199**, 323-330, doi:10.1006/viro.1994.1130 (1994).

31 Lyons, D. M. & Lauring, A. S. Evidence for the Selective Basis of Transition-to-Transversion Substitution Bias in Two RNA Viruses. *Molecular Biology and Evolution* **34**, 3205-3215, doi:10.1093/molbev/msx251 (2017).

32 Stoltzfus, A. & Norris, R. W. On the Causes of Evolutionary Transition:Transversion Bias. *Molecular Biology and Evolution* **33**, 595-602, doi:10.1093/molbev/msv274 (2016).

33 Rosenberg, M. S. Patterns of Transitional Mutation Biases Within and Among Mammalian Genomes. *Molecular Biology and Evolution* **20**, 988-993, doi:10.1093/molbev/msg113 (2003).
34 Chan, K.-F., Koukouravas, S., Yeo, J. Y., Koh, D. W.-S. & Gan, S. K.-E. Probability of Change in Life: amino acid changes in single nucleotide substitutions. *bioRxiv*, 729756, doi:10.1101/729756 (2019).
35 Eggington, J. M., Greene, T. & Bass, B. L. Predicting sites of ADAR editing in double-stranded RNA. *Nature Communications* **2**, 319, doi:10.1038/ncomms1324 (2011).
36 Tian, N. *et al.* A structural determinant required for RNA editing. *Nucleic Acids Research* **39**, 5669-5681, doi:10.1093/nar/gkr144 (2011).
37 Dawson, T. R., Sansam, C. L. & Emeson, R. B. Structure and Sequence Determinants Required for the RNA Editing of ADAR2 Substrates. *Journal of Biological Chemistry* **279**, 4941-4951, doi:10.1074/jbc.M310068200 (2004).
38 Bass, B. L. & Weintraub, H. An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* **55**, 1089-1098, doi:10.1016/0092-8674(88)90253-X (1988).
39 Nishikura, K. Functions and Regulation of RNA Editing by ADAR Deaminases. *Annual Review of Biochemistry* **79**, 321-349, doi:10.1146/annurev-biochem-060208-105251 (2010).
40 Clerzius, G. *et al.* ADAR1 Interacts with PKR during Human Immunodeficiency Virus Infection of Lymphocytes and Contributes to Viral Replication. *Journal of Virology* **83**, 10119-10128, doi:10.1128/jvi.02457-08 (2009).
41 Doria, M., Neri, F., Gallo, A., Farace, M. G. & Michienzi, A. Editing of HIV-1 RNA by the double-stranded RNA deaminase ADAR1 stimulates viral infection. *Nucleic Acids Research* **37**, 5848-5858, doi:10.1093/nar/gkp604 (2009).
42 Gélinas, J.-F., Clerzius, G., Shaw, E. & Gatignol, A. Enhancement of Replication of RNA Viruses by ADAR1 via RNA Editing and Inhibition of RNA-Activated Protein Kinase. *Journal of Virology* **85**, 8460-8466, doi:10.1128/jvi.00240-11 (2011).
43 Weiden, M. D. *et al.* Adenosine Deaminase Acting on RNA-1 (ADAR1) Inhibits HIV-1 Replication in Human Alveolar Macrophages. *PLOS ONE* **9**, e108476, doi:10.1371/journal.pone.0108476 (2014).
44 Lathouwers, E. *et al.* HIV-1 resistance rarely observed in subjects using darunavir once-daily regimens across clinical studies. *HIV Clinical Trials* **18**, 196-204, doi:10.1080/15284336.2017.1387690 (2017).
45 Titanji, B. K., Pillay, D. & Jolly, C. Combination antiretroviral therapy and cell–cell spread of wild-type and drug-resistant human immunodeficiency virus-1. *Journal of General Virology* **98**, 821-834, doi:10.1099/jgv.0.000728 (2017).
46 Sanchez-Merino, V., Farrow, M. A., Brewster, F., Somasundaran, M. & Luzuriaga, K. Identification and Characterization of HIV-1 CD8+ T Cell Escape Variants with Impaired Fitness. *The Journal of Infectious Diseases* **197**, 300-308, doi:10.1086/524845 (2008).
47 Sanchez-Merino, V., Nie, S. & Luzuriaga, K. HIV-1-Specific CD8+ T Cell Responses and Viral Evolution in Women and Infants. *The Journal of Immunology* **175**, 6976-6986, doi:10.4049/jimmunol.175.10.6976 (2005).
48 Mohamed, S. *et al.* Comparison of ultra-deep versus Sanger sequencing detection of minority mutations on the HIV-1 drug resistance interpretations

after virological failure. *AIDS* **28**, 1315-1324, doi:10.1097/qad.0000000000000267 (2014).

49 Parkin, N. T., Gupta, S., Chappey, C. & Petropoulos, C. J. The K101P and K103R/V179D Mutations in Human Immunodeficiency Virus Type 1 Reverse Transcriptase Confer Resistance to Nonnucleoside Reverse Transcriptase Inhibitors. *Antimicrobial Agents and Chemotherapy* **50**, 351-354, doi:10.1128/aac.50.1.351-354.2006 (2006).

50 De Iaco, A. & Luban, J. Cyclophilin A promotes HIV-1 reverse transcription but its effect on transduction correlates best with its effect on nuclear entry of viral cDNA. *Retrovirology* **11**, 11, doi:10.1186/1742-4690-11-11 (2014).

51 Banke, S., Lillemark, M. R., Gerstoft, J., Obel, N. & Jørgensen, L. B. Positive Selection Pressure Introduces Secondary Mutations at Gag Cleavage Sites in Human Immunodeficiency Virus Type 1 Harboring Major Protease Resistance Mutations. *Journal of Virology* **83**, 8916-8924, doi:10.1128/jvi.00003-09 (2009).

52 Fisher, T. S., Darden, T. & Prasad, V. R. Substitutions at Phe61 in the β3-β4 Hairpin of HIV-1 Reverse Transcriptase Reveal a Role for the Fingers Subdomain in Strand Displacement DNA Synthesis. *J Mol Biol* **325**, 443-459, doi:10.1016/S0022-2836(02)01225-1 (2003).

53 Ceccherini-Silberstein, F. *et al.* High Sequence Conservation of Human Immunodeficiency Virus Type 1 Reverse Transcriptase under Drug Pressure despite the Continuous Appearance of Mutations. *Journal of Virology* **79**, 10718-10729, doi:10.1128/jvi.79.16.10718-10729.2005 (2005).

54 Auwerx, J. *et al.* The N137 and P140 amino acids in the p51 and the P95 amino acid in the p66 subunit of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase are instrumental to maintain catalytic activity and to design new classes of anti-HIV-1 drugs. *FEBS Letters* **579**, 2294-2300, doi:10.1016/j.febslet.2005.02.077 (2005).

55 Santoro, M. M. & Perno, C. F. HIV-1 Genetic Variability and Clinical Implications. *ISRN Microbiol* **2013**, 481314-481314, doi:10.1155/2013/481314 (2013).

56 Chung, S. *et al.* Examining the role of the HIV-1 reverse transcriptase p51 subunit in positioning and hydrolysis of RNA/DNA hybrids. *Journal of Biological Chemistry* **288**, 16177-16184 (2013).

57 Betancor, G. *et al.* Effects of HIV-1 reverse transcriptase connection subdomain mutations on polypurine tract removal and initiation of (+)-strand DNA synthesis. *Nucleic Acids Research* **43**, 2259-2270 (2015).

58 Hu, W.-S. & Temin, H. M. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proceedings of the National Academy of Sciences* **87**, 1556-1560 (1990).

59 Yuste, E., Sánchez-Palomino, S., Casado, C., Domingo, E. & López-Galíndez, C. Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. *Journal of Virology* **73**, 2745-2751 (1999).

60 Ormond, L. *et al.* The combined effect of oseltamivir and favipiravir on influenza A virus evolution. *Genome biology and evolution* **9**, 1913-1924 (2017).

61 Summers, J. & Litwin, S. Examining The Theory of Error Catastrophe. *Journal of Virology* **80**, 20-26, doi:10.1128/jvi.80.1.20-26.2006 (2006).

62 Bull, J. J., Sanjuán, R. & Wilke, C. O. Theory of Lethal Mutagenesis for Viruses. *Journal of Virology* **81**, 2930-2939, doi:10.1128/jvi.01624-06 (2007).
63 Drake, J. W. & Holland, J. J. Mutation rates among RNA viruses. *Proceedings of the National Academy of Sciences* **96**, 13910-13913, doi:10.1073/pnas.96.24.13910 (1999).
64 Dapp, M. J., Clouser, C. L., Patterson, S. & Mansky, L. M. 5-Azacytidine Can Induce Lethal Mutagenesis in Human Immunodeficiency Virus Type 1. *Journal of Virology* **83**, 11950-11958, doi:10.1128/jvi.01406-09 (2009).
65 Harris, K. S., Brabant, W., Styrchak, S., Gall, A. & Daifuku, R. KP-1212/1461, a nucleoside designed for the treatment of HIV by viral mutagenesis. *Antiviral Research* **67**, 1-9, doi:10.1016/j.antiviral.2005.03.004 (2005).
66 Loeb, L. A. *et al.* Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proceedings of the National Academy of Sciences* **96**, 1492-1497, doi:10.1073/pnas.96.4.1492 (1999).
67 Chiang, Ron Z.-H., Gan, Samuel K.-E. & Su, Chinh T.-T. A computational study for rational HIV-1 non-nucleoside reverse transcriptase inhibitor selection and the discovery of novel allosteric pockets for inhibitor design. *Bioscience Reports* **38**, BSR20171113, doi:10.1042/bsr20171113 (2018).
68 Su, C. T.-T., Ling, W.-L., Lua, W.-H., Haw, Y.-X. & Gan, S. K.-E. Structural analyses of 2015-updated drug-resistant mutations in HIV-1 protease: an implication of protease inhibitor cross-resistance. *BMC Bioinformatics* **17**, 500, doi:10.1186/s12859-016-1372-3 (2016).
69 Naldini, L. *et al.* In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**, 263-267 (1996).
70 Chan, W. T., Verma, Chandra S., Lane, David P. & Gan, Samuel K. E. A comparison and optimization of methods and factors affecting the transformation of Escherichia coli. *Bioscience Reports* **33**, 931-937, doi:10.1042/BSR20130098 (2013).
71 Ling, W.-L. *et al.* Effect of VH–VL Families in Pertuzumab and Trastuzumab Recombinant Production, Her2 and FcγIIA Binding. *Frontiers in Immunology* **9**, doi:10.3389/fimmu.2018.00469 (2018).
72 Lua, W.-H. *et al.* The effects of Antibody Engineering CH and CL in Trastuzumab and Pertuzumab Recombinant Models: Impact on antibody production and antigen-binding. *Scientific Reports* **8**, 718 (2018).
73 Lua, W.-H. *et al.* Role of the IgE variable heavy chain in FcεRIα and superantigen binding in allergy and immunotherapy. *Journal of Allergy and Clinical Immunology* (2019).
74 Sim, J.-Z., Nguyen, P.-V., Lee, H.-K. & Gan, S. K.-E. GelApp: Mobile gel electrophoresis analyser. *Nature Methods Application Notes* (2015).
75 Poh, J. J. & Gan, S. K. E. The Determination of Factors involved in Column-Based Nucleic Acid Extraction and Purification. *Journal of Bioprocessing & Biotechniques* **04**, doi:10.4172/2155-9821.1000157 (2014).
76 Sim, J. Z., Nguyen, P. V., Zang, Y. & Gan, S. K. E. DNA2App: Mobile sequence analyser. *Scientific Phone Apps and Mobile Devices* **2**, 2, doi:10.1186/s41070-016-0004-7 (2016).
77 LosAlamos. HIV Sequence Database.