RESEARCH

ADAR and APOBEC editing signatures in viral RNA during acute-phase Innate Immune responses of the host-parasite relationship to *Flaviviruses*

Robyn A Lindley[1,2] and Edward J Steele[3,4]

[1]GMDxCo Pty Ltd, Melbourne Vic, AUSTRALIA; [2]Department of Clinical Pathology, Faculty of Medicine, Dentistry & Health Sciences, University of Melbourne, Vic, AUSTRALIA
[3]CYO'Connor ERADE Village Foundation, 24 Genomics Rise, Piara Waters, 6112, Perth, AUSTRALIA;
[4]Melville Analytics Pty Ltd, Melbourne, Vic, AUSTRALIA

Corresponding authors' email: ejsteele@cyo.edu.au and robyn.lindley@unimelb.edu.au

ABSTRACT

The origin of variability in positive single-stranded RNA *Flaviviruses* is poorly understood. Is the origin of the high rate of spontaneous mutations arising early in *Flavivirus* infections due solely to error-prone viral replication? Alternatively, are these mutations due to viral replicase incorporation of high frequency RNA deamination editing events? The data analyzed here strongly supports the second explanation. The viral genomes analysed include: a) acute phase Zika virus (ZIKV) genomes associated with microcephaly; b) hepatitis C virus (HCV) genomes from both acute and chronic infected hepatitis patients; and c) hepatitis B virus (HBV) genomes of patients infected from the early/acute phase of a large nosocomial outbreak. RNA mutations at motifs for APOBEC and ADAR deaminases were analyzed within the codon context structure, used previously for targeted somatic mutation (TSM) analyses of cancer exomes. The results show that transition mutations targeting MC3 nucleotide sites (ie. the third nucleotide position within the structure of the Mutated Codon read 5' to 3'), and within known RNA editing motifs for APOBEC1, APOBEC3A and ADAR1/2 account for the majority of the mutations for ZIKV, HCV, as well as for HBV. The results also imply that drug therapy strategies might profitably focus on the RNA-based secondary structure of potential "druggable pockets" within apparently conserved regions of amino acid sequence within variant viral strains. Such sites could encode subtle changes in RNA secondary structures as potential vulnerable target regions in pathogenic variants. For ZIKV, such sites could include the hypermutable MC3-deaminase targeted sites.

KEYWORDS: ZIKV, HCV, *Flaviviruses*, ADAR/APOBEC RNA Editing, RNA mutations, Drug Discovery

Citation: Lindley RA and Steele EJ (2018) ADAR and APOBEC editing signatures in viral RNA during acute-phase Innate Immune responses of the host-parasite relationship to *Flaviviruses*. Research Reports 2:e1-e22. doi:10.9777/rr.2018.10325.

INTRODUCTION

Zika virus and other recognized viral pathogens such as HCV, HBV and HIV-1 exact huge tolls on the health of hundreds of millions around the globe. There are currently no treatment options available for ZIKV infection, and the host-parasite relationship of hepatitis C virus-induced liver disease still requires much attention. Increasing our understanding of virus mutation origins through nucleotide sequence target specificity might assist in the identification of new antiviral strategies.

The impact of established innate anti-viral deaminase mutations C-to-U and A-to-I in RNA (and DNA) editing on Zika virus genomes is compared here with similar anti-viral deaminations in another *Flavivirus* of medical significance, acute and chronic hepatitis C virus (HCV), and with the dsDNA/RT para-retrovirus causing liver disease, hepatitis B virus (HBV). Increasing our understanding of virus mutation origins through nucleotide sequence target specificity will hopefully assist in therapeutic intervention in drug discovery for both druggable pockets in the viruses themselves (Cox et al. 2015) as well as the development of oligonucleotide "deaminase blockers" in association with the crystal structures of the anti-viral genome deaminases (Bohn et al. 2013; Logue et al. 2014; Shi et al. 2015; Armitage et al. 2014). There is also now a renewed focus on the role of the APOBEC3 related deaminases (AID, APOBEC1, APOBEC3A –> APOBEC3H) which includes the well-studied anti-retroviral APOBEC3G deaminase, as direct DNA deamination viral restriction factors (Conticello 2008; Refsland and Harris 2013; Armitage et al 2014; Monajemi et al 2012; Barrett et al 2014; Naraiza et al 2009; Logue et al 2014).

Single-stranded RNA viruses display great intrastrain genetic diversity with higher mutation rates than single stranded (ss) or double stranded (ds)DNA viruses (Sanjuan et al. 2010; Sanjuan 2012; Sanjuan and Domingo-Calap 2016). Their median mutation rates of 10^{-5}-10^{-3} substitutions per nucleotide per cell infection are comparable to the hypermutable diploid RNA retroviruses such as HIV-1 and are clear strategic factors in viral immune evasion and resistance to anti-viral drugs. However the viruses themselves are non-cytolytic and benign, with the initial infection often being silent and asymptomatic. The main collateral damage to human health occurs later, due to inflammatory cellular tissue invasion by the host's adaptive immune response as chronic status is established in those subjects who fail to naturally clear the infection.

As with all viral infections, the concept of "quasi species" applies (Eigen and Schuster 1979). That is, within each infected host a heterogeneous yet genetically related set of viruses are abundant – establishing a type of "mutant swarm" (Andino and Domingo 2015). Thus "..quasispecies theory has encouraged viewing and examining viruses as complex mutant spectra, and not merely as simple genetic entities that can be described with a defined nucleotide sequence, as believed only a few decades ago" (Andino and Domingo 2015). Therefore, the single sequence established from a plasma sample, is usually the most frequent or consensus sequence of a genetically diverse virion population. Viral diversity can be described as a series of bottleneck transmissions: once in the new host, rapid replication and diversification produces the mutant swarm, then on infection of the next new host a sieving process produces a new bottleneck. The virion and viral gene pool within an infected individual shows variability in virulence, in terms of infectivity, speed of uncoating and replication, burst size, proportion of empty and defective virions and thus efficacy in transmission to new hosts.

It is also clear from deep sequencing studies involving single genome sequencing (SGS) that transmitted founder (T/F) viruses establish early in the acute phase of a bottleneck (Keele et al. 2008; Ribiero et al. 2012; Li et al. 2012; Stoddard et al. 2015). Sequences within a virus lineage display a

distinct "star-like phylogeny" of mutations that coalesce to one or a few unambiguous T/F viral genomes. Despite this precision in T/F identification, the "viral swarm" is the effective inoculum. This is presumably to maximize gene complementation effects in the newly infected cell that can, for example, even maintain defective viral genomes carrying stop codons (Combe et al. 2015; Aaskov et al. 2006). Thus "viral aggregations", often enwrapped in a membrane vesicle, appear to be the actual infective dose rather than a productive infection from just a single virion entering a target cell (Chen et al. 2015; Combe et al. 2015).

In the discipline of Virology it is widely assumed that the polymerase enzymes ("replicases") responsible for viral RNA replication, the RNA-dependent RNA polymerases (RdRP), are intrinsically error-prone as they do not possess proof reading activity. This high replicative error rate provides the mutations necessary for rapid virus evolution and host adaptation regularly noted *in vivo* for many years (Holland et al 1982) - allowing viral evasion strategies against either clinical drug therapies or host Innate Immunity and Adaptive Immune responses. What is striking about this widely held assumption is that it is not strictly supported by direct experimental evidence (Freistadt et al. 2007; Aggarwal et al. 2010; (reviewed in te Velthuis 2014). However, when detailed biochemical analyses have been performed to measure the actual RdRP enzyme fidelity *in vitro,* as in the case of Poliovirus (Freistadt et al. 2007) and Influenza A virus (Aggarwal et al. 2010), the RdRP replicases actually display relatively high fidelity for polymerases without proof reading functions. In particular, the polymerases display very low error rates for transversion mis-incorporation mutations. Further, for most transition mutations, as shown by Powdrill et al (2011), the distinct *in vitro* biochemical consequences of the well known high frequency G•U mispair misincorporations (by RdRP) can theoretically cause high rates of C-to-U transitions. Following further viral RNA replication, these primary C>U transitions may then produce all the other transitions also at reasonable high frequency (G>A, A>G, T>C). This is theoretical because the imbalances of ribonucleoside triphosphate (NTPs) precursor pools *in vivo* in infected patients is an unknown variable. Indeed Powdrill et al (2011) conclude: "A point of caution is that biochemical measurements are restricted to a few sites, and error rates per genome replication can only be extrapolated and not accurately determined from this data. Moreover, effective concentrations and variations of intracellularly available nucleotide pools at the site of viral replication are unknown".

What is not theoretical is the extensive literature documenting the significant impact *in vivo* on viral genomes of the APOBEC (C-to-U) and ADAR (A-to-I) deaminases which demonstrably alter viral RNA (and DNA) sequences, often at high frequency (Murphy et al 1991; Casey and Gerin 1995; Haijar and Linial 1995; Kim et al 1996; Bourara et al 2000; Bass 2002; Martinez and Melero 2002; Harris et al 2003; Mangeat et al 2003; Taylor et al 2005; Xu et al 2007; Zahn et al 2007; Conticello 2008; Koch and Blum 2008; Gonzalez et al 2009; Vartanian et al 2010; Samuel 2011; Monajemi et al 2012; Refsland and Harris 2013; Armitage et al 2014). Such deamination events on RNA substrates for APOBEC1, APOBEC3A and ADAR enzymes have also been documented at similar agreed motifs in both unpaired RNA loops and in double stranded RNA secondary structures (Sharma et al 2015, 2016; Rosenberg et al 2011; Bass 2002; Eifler et al 2013).

The RdRP polymerases thus join with other relatively high fidelity RNA template-based polymerases conserving essential genomic functions, such as the DNA polymerase- γ which is responsible for the faithful replication of the mitochondrial genome via an RNA intermediate step (Murakami et al.2003). However there are also clearly other error-prone polymerases involved in an organism's adaptive responses to environmental stress where they serve a beneficial

function. These include the Y family DNA polymerase eta (η) involved in both translesion DNA repair of thymine-dimers and in antibody diversification (Zeng et al 2001; Goodman 2002) and which also displays RNA template-based reverse transcriptase activity (Franklin et al. 2004; Steele and Lindley 2017); and there are other related error-prone DNA-dependent polymerases involved in the bacterial adaptive SOS "hypermutation" response (Radman 2005; Bridges 2005; Tippin et al 2004).

It is therefore pertinent to ask: Just what are the causes of the widely observed high viral genomic variability *in vivo*? Such causes should be viewed within the framework of the host-parasite relationship. No doubt they involve both host factors causing viral genome diversity, as well as virus co-factors promoting such diversity (e.g. the encapsulation of the host retroviral restriction factor APOBEC3G deaminase protein within HIV virions, (reviewed in Conticello 2008; Refsland and Harris 2013).

Viral mutation rates can be greatly elevated by APOBEC-mediated C-to-U DNA deamination (restriction) events in both retroviral (HIV-1) and para-retroviral (HBV) genomes (Bourara et al. 2000; Harris et al. 2003; Lecossier et al. 2003; Mangeat et al. 2003; Chiu et al. 2005; Xu et al. 2007; Kock and Blum 2008; Gonzalez et al. 2009; Vartanian et al. 2010; Sanjuan and Domingo-Calap 2016). For retroviruses and pure RNA viruses numerous examples of RNA editing via the A-to-I dsRNA ADAR1/2 deaminases have been reported (Murphy et al. 1991; Hajjar and Linial 1995; Casey and Gerin 1995; Kim et al. 1996; Bass 2002; Martinez and Melero 2002; Taylor et al. 2005; Zahn et al. 2007; Samuel 2011). These host innate immune responses can pose an obstacle to viral replication as many of the mutations, in finely balanced information-dense viral genomes, will produce mutant variants crippled in their speed of replication and infectivity, or prevent either replication or transmission by being outright lethal mutations. In other cases A-to-I editing can enhance viral fitness (Samuel 2011). In site-directed mutagenesis experiments studying the effect of such randomized mutations on viral genomes (Sanjuan et al. 2004; Sanjuan 2010) up to 40% are clear lethals, and the consensus opinion is that upwards of 90% of random single-nucleotide mutations may have some effect on reducing replicative fitness (reviewed in Lauring et al. 2012). Thus, purifying selection acts on the great majority of random viral mutations which is likely to be reflected in the mutation signatures observed in the surviving virus population detected and sequenced.

However, despite all these studies, the origin of mutations in RNA viruses particularly the *Flaviviruses* like HCV and ZIKV, or RT-viruses (reverse transcriptase-dependent) like HBV and HIV, needs to be understood much better than at present. Considerable confusion exists in the literature on the origins of these "spontaneous mutations", particularly for the RNA viruses. In the early acute phase such "mutant swarms" are often referred to as having arisen "essentially randomly" (Stoddard et al. 2015) assumed to be by the error-prone replication by RdRP replicases (Powdrill et al. 2010; Li et al. 2012; Ribiero et al. 2012). This assumption is actually a theoretical one, as outlined above. It is thus surprising because the assumption both ignores direct experimental demonstrations of the relatively high fidelity of RdRP RNA replicases (Freistadt et al. 2007; Aggarwal et a 2010) as well as the well-known mutagenic role of the host anti-viral activities via APOBEC and ADAR deaminases already discussed here at length.

The main focus of this paper is concerned with the origins of mutations in the positive single-stranded RNA ZIKV virus, as an effective drug therapy cure for this virus is in great demand. By way of comparison, the likely origin of mutations arising in the HCV RNA virus (acute and chronic), as well as the para-retrovirus HBV were also analysed. For acute phase ZIKV, the group of ZIKV variants detected early in the 2013-2015 South American

Zika epidemic associated with microcephaly (Mlakar et al. 2016) were analysed. For HCV the causes of the variability will be sought in the Transmitter/Founder (T/F) virus studies in both acute and chronic HCV (HCVa and HCVc) infected patients (Ribiero et al. 2012; Stoddard et al. 2015). It is the T/F virus population which is the target of early attempts via the body's innate or adaptive immune mechanisms to clear the infection (either through anti-viral APOBEC/ADAR deaminations or antibody viral neutralization or killing of the virus infected cell by CD8+ cytotoxic T cells). It is clear in the HCV T/F studies that in the period to 28 days post infection there is little if any adaptive immune response (Stoddard et al. 2015) as expected from other studies (Sung et al. 2014; Thimme et al. 2012; Shin et al. 2011). Presumably this delay in adaptive immunity also applies to the acute phase of ZIKV infection in the cases examined in Mlakar et al (2016) which are analysed here.

It is hypothesized that the new mutations arising *de novo* in RNA and RNA-based viruses arise in a highly non-random manner. Further, it is postulated that it is the site-targeted nucleotides within the context of the codon structure governing the preferential C-to-U and A-to-I RNA editing deaminase activity of host APOBEC and ADAR proteins which generates these mutations, and which are then expected to be faithfully propagated by RdRP replicase complexes in the cytosolic membranous web (Thimme et al. 2012). The resulting new variants are predicted to play a key role in the generation of new viral strains after selection.

MATERIALS AND METHODS

Viral genome DNA sequence source

All DNA sequences analysed were downloaded from GenBank following the Accession numbers reported in each source publication. It should be noted that the ZIKV and HBV datasets are mutation patterns derived from the dominant-consensus plasma sequence in each human subject; for the HCV datasets, both acute and chronic, the mutation patterns derive from the quasi-species in the plasma from individual subjects by deep SGS sequencing.

The FASTA formats were uploaded into and then aligned using Clustal Omega (http://www.egi.ac.uk). In an alignment the difference from the reference/consensus where different clones share the same change is only scored once for that position. The determination of the reference/consensus differed for each data set and these criteria are indicated. It is understood that T = U in RNA genomes.

ZIKV whole RNA genome mutation data

Supplementary Data: Table S1 displays the resulting codon-context data of the mutations found in ZIKV whole RNA genomes reportedly associated with the South American Zika virus epidemic 2013-2015 (Mlaker et al. 2016). The GenBank Accession numbers of the ZIKV genome sequences isolated in French Polynesia (2013) and Brazil (2015, 2016) and showing association with Microcephaly are provided in Mlakar et al (2016). These sequences, and other earlier isolates reported on for the phylogenetic tree shown in Figure 4 in Mlakar et al. (2016) were downloaded from GenBank and the FASTA formats aligned using Clustal Omega (http://www.egi.ac.uk). Given that the 2013-2015 strains circulating in French Polynesia and Brazil appear to have emerged from the Asian lineage (Faye et al 2014), the reference sequence was the consensus of the 2007 (Micronesia) and 2010 (Cambodia) outbreaks, but where the 2007 Micronesia isolate (EU545988) differed the default reference was to the more recent 2010 Cambodian strain (JN860885). Location of CDS Start and CDS End sites in the polyprotein was made by reference to Uganda (1947) Zika virus strain MR766 (httpp://www.viprbrc.org/brc/viprStrainDetails.spg?ncbiAccession=AY632535&decorator=flavi). In Table S1, a difference from the reference sequence that is common to the whole GROUP of eight circulating viruses in the South American epidemic is referred

to as "Group" and a single nucleotide variation within this group is labelled a "Variant" as the reference nucleotide is defined by the "Group" consensus.

A complete list of all of the single nucleotide variations, their codon position, position within the codon, protein product locations, synonymous/non-synonymous category, type of amino acid change, whether transition or transversion is as shown in Table S1. This table also shows the mutated position in the codon as 1st (MC1), 2nd (MC2) or 3rd (MC3) according to mutated codon-context method of Lindley (2013) where the target codon and the 5-prime (5') and 3-prime (3') flank codons of the reference sequence are annotated (read 5' to 3'). Sequence Manipulation Suite (version 2) was used to determine nucleotide locations and statistics, particularly codon usage and base composition at MC3 for 2010 Cambodian strain (JN860885). (http://www.bioinformatics.org/sms2/index.html).

Acute HCV (HCVacute) whole RNA genome mutation data

Supplementary Data: Table S1 displays the codon-context data of HCV whole RNA genomes associated with the early stages of acute Hepatitis, genotype 4a, (up to day 28 post infection) isolated and sequenced from a single patient (105431) reported on by Stoddard et al. (2015). The HCV viral RNA genomes, a quasi species set of "evolving" sequences in a single patient (105431) from the day of known infection until about day 28 were isolated from plasma and subjected to single-genome sequencing (SGS). For details refer to the data summaries in Figures 1 and 2, and Table 1 in Stoddard et al. (2015). All these patients, including 105431, were weekly plasma donors and the time of infection could be accurately determined. The GenBank Accession numbers for the various full length reference sequences of the two founder 105431 transmitted viruses (TF1, TF2) plus all the various sequence sets of the 5' and 3' amplicons and cloned portions are: TF1, KM043282-full length genome; TF2, KM043284-full length genome; KP666553-KP666592 Clone 105431 5' IRES + 5' CDS Amplicons -668bp (IRES, internal ribosome entry site); KP666630-KP666688 Clone 105431 5h (= 5' half see Fig 2 Stoddard et al 2015) Amplicons - 4906bp; KP668636-KP668678 Clone 105431 3h (= 3' half see Fig 2 Stoddard et al 2015) Amplicons – 4597bp; and, KP666311-KP666378 Clone 105431 3'CDS + 3' UTR (see Figure 2 in Stoddard et al. 2015) Amplicons - 381bp.

Table 1. Expected Proportions (%) of all Potential Random Mutations [†].

		Mutated Codon			Type of Mutation		
		MC1	MC2	MC3	Total	Transitions	Transversions
ZIKV	Sum of Proportions	33.3	33.3	33.3	100	33.3	66.7
	Synonymous	1.7	0	21.7	23.4	10.9	12.4
	Non-Synonymous	29.4	32.3	10.6	72.2	21.5	50.7
	Stop	2.3	1	1.1	4.4	0.9	3.5
HCVa	Sum of Proportions	33.3	33.3	33.6	100.2	33.4	66.9
	Synonymous	1.6	0	23.8	25.3	11.3	14.1
	Non-Synonymous	30.4	32.4	8.5	71.3	21.3	50
	Stop	1.5	1	1.3	3.7	0.9	2.8

[†] No Selection. Adjusted for codon usage rounded to one decimal place (Supp data Table S2)

Comparative RNA sequence analyses across these genomic sequences were done using Clustal Omega multiple sequence alignment. The TF1 and TF2 reference sequences guided the analysis and all sequence differences were by reference to TF1. Protein product boundaries in the polyprotein were determined by reference to the H77 genome accession NC_004102 http://hcv.lanl.gov/content/sequence/HCV/MAP/landmark.html. The tabulated data, including details of annotations, are included in Table S1.

A complete list of all the nucleotide variations their codon position, position within the codon, protein product locations, synonymous/non-synonymous category, type of amino acid change, whether transition or transversion is as shown in Table S1. Sequence Manipulation Suite (version 2) was used to determine nucleotide locations and statistics, particularly codon usage and base composition at MC3 sites for the acute phase 105431 TF1 HCV strain.http://www.bioinformatics.org/sms2/index.html.

Chronic HCV (HCVchronic) 5′ half genome RNA mutation data

HCV viral RNA sequences "evolving" in a single patient (ARJA6267) suffering chronic HCV (HCVc) viral hepatitis were analysed for codon-contexted sequence differences within the quasi-species population of virions isolated from the patient's plasma. Only the 5′ half of the genomes were subject to single-genome sequencing. The 4875 bp RNA sequences were downloaded from GenBank (KP668412-KP668450) and comparative sequence analysis conducted as before using Clustal Omega multiple sequence alignment at (www.ebi.ac.uk/Tools/msa/clustalo/)

A complete list of all of the nucleotide variations, their codon position, position within the codon, protein product locations, synonymous/non-synonymous category, type of amino acid change, whether transition or transversion is as shown in Table S1. Sequence Manipulation Suite (version 2) was used to determine nucleotide locations, codon usage and base composition at MC3 sites http://www.bioinformatics.org/sms2/index.html

HBV whole DNA genome mutation data

The full length DNA genomic sequences of HBV strains (genotype A, subgenotype A2) isolated from 26 matched "control" patients during 2003-2008 in the Limburg region of Belgium with presumptive association with acute hepatitis reported on by Pourkarim et al. (2009) were downloaded from GenBank. The Accession Numbers for the 26 "C" isolates are included in Figure 1 of Pourkarim et al. (2007). (Note that C56 at EU859952 was included incorrectly in this GenBank filed set and is a clear outlier; it is not included in the phylogenetic analyses of Figure 1 of Pourkarim et al. 2009 and thus not included in the comparative sequence analysis included in this study.) The sequences of these full length DNA genomes were compared using Clustal Omega multiple sequence alignment at (www.ebi.ac.uk/Tools/msa/clustalo/). Nucleotide position numbering was by reference to the isolate from patient C126, and all overlapping reading frames of the Polymerase, Large S protein, X peptide and Core protein determined by reference to positions filed in the GenBank submissions. The tabulation and annotation of all of the nucleotide differences from the consensus of the 26 sequences is included in Table S1.

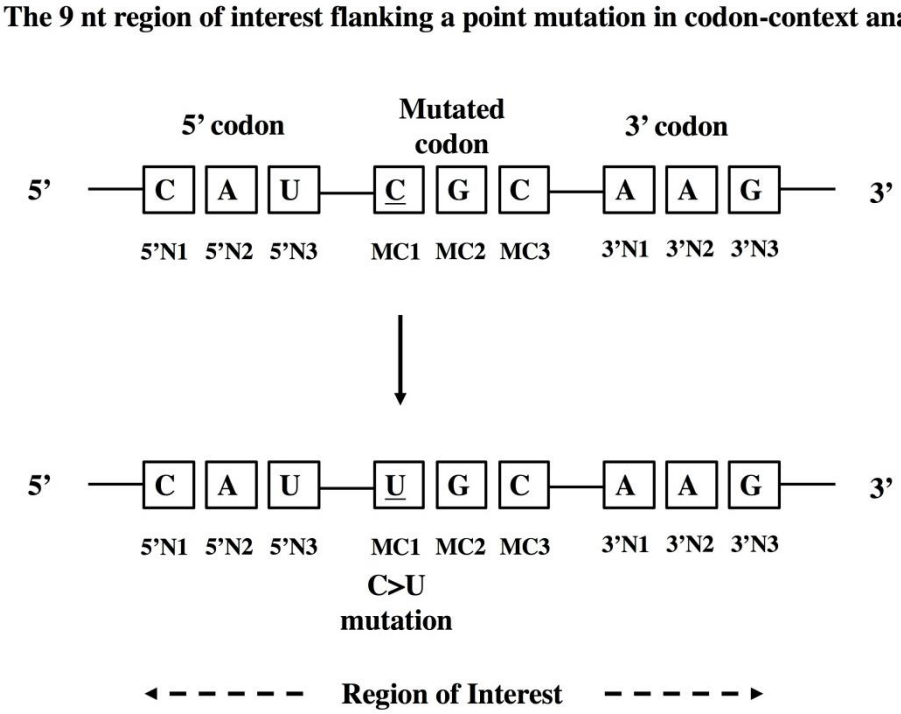

Figure 1. The Codon-Context TSM Analytical Method. Adapted from Figure 1 in Lindley (2013). The defined 9N region of interest flanking a point mutation in a viral RNA exome sequence is shown. An example of a single point mutation in the region of interest and the annotations used to identify each of the nucleotides for this analysis. The region of interest surrounding the mutated codon (MC) includes the base composition of the flanking 5′ and 3′ codons. The nucleotide sites in the MC are annotated as MC-1, MC-2, and MC-3, respectively (read 5′ to 3′). A C-to-U conversion at an APOBEC3A AUCC motif is shown on the + RNA strand which after RNA synthesis will register as a G-to-A point mutation on the - RNA strand (and vice versa).

Codon-contexted mutation profiling of ZIKV, HCV and HBV mutation data

A summary of the basic mutation patterns extracted from the viral data sets (Table S1) were tabulated and the results (O, observed) are shown in Table 2 for ZIKV (acute), HCVacute, HCVchronic, and HBV (acute) non-overlapping reading frames (Polymerase, X protein, Core). Table 2 shows the mutated position in the mutated codon (MC) as either 1st (MC1), 2nd (MC2) or 3rd (MC3) nucleotide, read 5′ to 3′. A breakdown of the relative numbers of synonymous and nonsynonymous mutations, and the number of transitions and transversions was also tabulated from the mutation data shown in Table S1, and included in Table 2. Observed values are compared to the Expected under a pure random model without selection (Table 1).

Table 2. Codon-contexted mutation pattern in Zika Virus (ZIKV), Hepatitis C Virus Acute (HCVa), Hepatitis C Chronic (HCVc) whole genome datasets, and Hepatitis B Virus (HBV) mutation dataset from non-overlapping reading frames (Polymerase, X protein, Core), where MC1, MC2 and MC3 refer to the location of each mutation within the structure of the mutated codon (MC, read 5-prime to 3-prime).

A. ZIKV		MC1	MC2	MC3	Total	Transitions	Transversions
No. Mutations	O	26	5	178	209	195	14
	E	70	70	70		70	139
	P-value	<0.001	<0.001	<0.001		<0.001	<0.001
Synonymous	O	11	0	177	188	175	13
	E	4	0	45	49	23	26
	P-value	<0.001	ns	<0.001	<0.001	<0.001	<0.05
Nonsynonymous	O	14	5	2	21	20	1
	E	61	68	22	151	45	106
	P-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
B. HCV Acute		**MC1**	**MC2**	**MC3**	**Total**	**Transitions**	**Transversions**
No. Mutations	O	15	14	64	93	83	10
	E	31	31	31		31	62
	P-value	<0.01	<0.01	<0.001		<0.001	<0.001
Synonymous	O	3	0	61	64	60	4
	E	2	0	22	24	11	13
	P-value	ns	ns	<0.001	<0.001	<0.001	<0.05
Nonsynonymous	O	12	14	3	29	23	6
	E	28	30	8	66	20	47
	P-value	<0.01	<0.01	ns	<0.001	ns	<0.001
C. HCV Chronic		**MC1**	**MC2**	**MC3**	**Total**	**Transitions**	**Transversions**
No. Mutations		58	36	240	334	298	36
Synonymous		8	0	234	242	232	10
Nonsynonymous		50	36	6	92	66	26
D. HBV No-Overlap		**MC1**	**MC2**	**MC3**	**Total**	**Transitions**	**Transversions**
No. Mutations		15	9	44	68	40	28
Synonymous		2	0	41	43	31	12
Nonsynonymous		13	9*	3	25	9*	16

* One of these is a STOP codon, TAG

O, Observed; E, expected calculated from proportions in Table 1; P-value for Chi-square 1 df.

ns, not significant

No formal Obs. v Exp. Chi-square statistics are presented for C. and D. first because the aim of this paper has been to ascertain the targeted mutational patterns in Flaviviruses in the early or acute phase of infection in the absence of adaptive B and T cell immune selection. i.e. during the Innate Immune phase only. Further, in the case of the HBV no-overlap this data represents a small fraction of the total genome and the mutation numbers, while trending similarly for MC3 mutation targeting, are on the small side, together with the complexity of the genome and replication cycle compared to ZIKV and HCV.

Identifying the deaminase source of MC3 Transitions in ZIKV and HCV

Given that the majority of mutations are synonymous transitions and these fall in MC3 sites, we next examined the sequence context motifs as substrates for likely APOBEC RNA deaminases. The only known genome-wide C-to-U RNA editors are APOBEC1 and APOBEC3A (Rosenberg et al. 2011; Sharma et al 2015; Sharma et al 2016). It is a given that since both positive (+) strand and complementary (-) strand RNA intermediates will exist *in vivo* then both C>U (and complement G>A events) at previously identified deaminase motifs are indicative of C-to-U deaminations subsequently preserved by viral RdRP replicase copying.

For APOBEC1 a number of high frequency motifs have been identified in transcriptome-wide analyses (Rosenberg et al. 2011) – expressed as a simple tri-nucleotide deaminase RNA motif these are A<u>C</u>A, A<u>C</u>U, U<u>C</u>A, U<u>C</u>U. It is known that 5′ and 3′ nucleotides 2 to 4 nt positions either side of the <u>C</u> target also play a role in C>U targeting (Rosenberg et al. 2011). For APOBEC3A, a similar type of targeting applies (Sharma et al 2016) which can also overlap with APOBEC1 targeting – expressed as a simple tri-nucleotide deaminase RNA motif these are U<u>C</u>G, U<u>C</u>A, C<u>C</u>G, C<u>C</u>A and a U can also be located on the 3′ side. In addition, 5′ and 3′ nucleotides 2 to 4 nt positions either side of the <u>C</u> target also play a role in C>U targeting (Sharma et al. 2016).

Using these motifs, the mutation data sets (Table S1) were queried to identify and tabulate the number of mutations coincident with these tri-nucleotide motifs for APOBEC1 and APOBEC3A. The results are shown in Table 3 for MC3 transitions using ZIKV acute RNA, and HCVacute, HCVchronic RNA whole genome data sets.

Genome-wide deamination events involving A-to-I RNA editing, revealed as A>G and U>C mutations, catalyzed by adenosine deaminases active on dsRNA substrates (ADAR1, ADAR2) universally target A-nucleotides usually with a 5′ A or U (= W) (Bass 2002). However it is clear that the base flipping process (Bass 2002; Kuttan and Bass 2012) to gain access to the A deaminase substrate in the context of a dsRNA A:U base pair probably also depends critically on surrounding A/U (W) base pairs, which appears to be the case in ADAR2 transcriptome-wide expression studies in yeast (Eifler et al. 2013). Thus A-targets flanked by Ws which are 2 nt removed either side are also common. For this reason, MC3 A-site transitions at W<u>A</u> /<u>U</u>W targets, and WS<u>A</u> /USW sites (where S = C or G) were identified. The tabulated results showing the number of mutations at these ADAR target sites are also shown in Table 3 for MC3 transitions using ZIKV acute RNA, and HCVacute and HCVchronic RNA whole genome data sets.

RESULTS

A new method of viral RNA mutation signature analysis is used here based on the codon-context of mutations termed ‘targeted somatic mutation’ (TSM, Lindley 2013; Lindley et al 2016). Recently, it has been shown that the pattern of mutations in cancer genomes caused by both AID/APOBEC and ADAR deaminases occurs in a codon-context fashion (Lindley 2013; Lindley et al 2016; Steele and Lindley 2017). Indeed using the TSM method it has been shown that the great majority of clinically significant human SNPs appear to have arisen by AID/APOBEC and ADAR deaminations leading to these germline SNPs (Lindley and Hall 2018). A plausible cause of the codon-context pattern may be due to A-site and C-site enzymatic deaminations on dsRNA and ssDNA moieties (as well as RNA: DNA hybrid substrates, Zheng et al 2017) in the context of stalled Transcription Bubbles (Steele and Lindley 2017). The TSM codon-context annotation method is shown in Figure 1 with the RNA editor APOBEC3A targeting a C-site motif as an example. The data analysed are displayed in the codon context tables in the excel book Table S1. This approach is now used to

Table 3. Number of MC3 C>U/G>A and A>G/U>C transition mutations occurring off motifs known to be associated with APOBEC1, APOBEC3A and ADAR1/2 RNA-targeting deaminase activity for ZIKV acute, HCV acute and chronic whole RNA genomes

Likely Deaminase	C-to-U Motif (+ve RNA Strand)	ZIKV	HCVa	HCVc	G-to-A Motif (-ve RNA Strand)	ZIKV	HCVa	HCVc
APOBEC3A/1*	UCA	5	3	7	UGA	1	1	2
APOBEC3A	ACA	8	0	4	UGU	1	0	1
	GCA	3	0	3	UGC	2	0	1
	CCG	4	2	8	CGG	3	5	4
	UCG	3	0	7	CGA	1	0	1
	ACG	6	5	6	CGU	0	2	2
	GCG	1	4	3	CGC	0	1	4
APOBEC1	CCA	6	0	8	UGG	1	0	5
	CCU	2	1	4	AGG	4	0	4
	UCU	1	1	2	AGA	7	1	1
	ACU	2	0	3	AGU	1	0	3
	GCU	2	1	8	AGC	4	0	4
	A-to-G Motif (+ve RNA Strand)				U-to-C Motif (-ve RNA Strand)			
ADAR1/2[t]	WA	18	4	13	UW	29	8	29
	WSA	6	4	5	USW	15	2	18
	AW	1	0	9	WU	6	2	9
	ASW	4	3	5	WSU	2	0	6
	Total	72	28	96		77	22	94

Proportion of ZIKV MC3 transitions off above motifs is (72+ 77)/195 = 76.4%
Proportion of HCVa MC3 transitions off above motifs is (28+22)/83 = 60.2%
Proportion of HCVc MC3 transitions off above motifs is (96+94)/298 = 63.8%
Edited nucleotide is underlined
*UCA/UGA is a motif targeted by both APOBEC1 and APOBEC3A
[t]W=A/U,S=G/S

establish whether or not the codon-context also influences the preferred target sites for *de novo* viral RNA mutations at C-sites and A-sites at known APOBEC and ADAR motifs. Given that virus genomes are known substrates for DNA and RNA deamination (C-to-U, A-to-I), the codon-context pattern of mutations in viruses of the positive strand RNA *Flavivirus* genus (ZIKV, HCV) and in HBV (a dsDNA Hepadnavirus which replicates via an RNA intermediate) were analysed. For ZIKV, HCV and HBV, the codon-context mutation pattern in genomes from the early acute infection phase (Mlakar et al. 2016; Stoddard et al. 2015; Pourkarim et al. 2009), and for the chronic phase in HCV (6 months to many years post infection, Stoddard et al. 2015) was tabulated.

The codon-contexted mutation profiles of ZIKV, HCV and HBV (in the non-overlapping reading frames) are shown in Table 2. The expected proportions based on a model of random mutation without any selection within ZIKV and the HCV acute genomes are shown in

Table 1. These expectations are dictated by the logic of the Genetic Code showing that transversion mutations are the overwhelming potential source of non-synonymous mutations, mainly at the first and second codon positions (amino acid changing). These expectations are for ZIKV and HCV viruses based on a pure random mutation model adjusted for codon usage across all sites without any selection (Table 1). This random model ignores the theoretical possibility of G•U mispair/misincorporation as discussed above and assumes optimum ribonucleoside triphosphate precursor pools *in vivo*.

Clearly the observed numbers of mutations (Table 2) in most categories show highly significant (P<0.01 at least) enhancements or deficits over what might be expected without the operation of some type of non-random effect or selection for viable virus.

The over-arching pattern in the observed mutation patterns in Table 2 is the significant reversal of the expected Transition:Transversion ratio (1:2) if all types of mutations were to be generated randomly which gives high Chi-square values for ZIKV and HCVacute (P<0.001), and reduced yet still significant reversals in the expected ratio in the non-overlapping reading frames of HBV (P<0.01). Among transitions, large significant deficits are observed for most categories of non-synonymous mutations (P<0.01 at least) suggesting selection for viable viral genomes. In the case of ZIKV the vast majority (93%) of all RNA mutations in these early South American isolates are transition mutations. For the HCVacute and HCVchronic datasets the proportion of transitions is also very high (≥89%). This is a striking enrichment. Similar high proportions (≥ 90%) of transition mutations were noted in the twenty treatment-naive HCV patients reported on by Powdrill et al (2011). However, the Powdrill et al study lacked information on the time when the plasma sample was taken after first report of HCV infection, as well as RNA sequence data (NS3) to enable codon-context assessment of that data.

Thus, the dominant feature in the ZIKV and HCVacute data (Table 2, A and B) is the enrichment for synonymous transitions at MC3. Upwards of 85% (ZIKV) and 66% (HCVa) of all mutations are MC3 synonymous transitions. In ZIKV, 90% of all mutations are synonymous. Diversity in terms of proportions of nonsynonymous mutations, particularly transversions is higher in HCVacute than ZIKV. In HCVchronic large increases in the proportions of nonsynonymous transitions and transversions are evident, and indicative of a greater increase in intrastrain diversity within the host. Yet even in the case of HCVchronic (Table 2C) the proportion of total mutations which are MC3 transitions is high at 70%.

The pattern of mutation is far more complex in HBV mutations (Table 2D). The potential nucleic acid substrates for mutation here are both RNA and DNA. The mutations for the regions of the genome with non-overlapping reading frames shows that the proportion of total mutations of MC3 transitions is 60.3%, which is comparable to HCVacute.

Using the mutation data and nucleotide context adjacent to each mutation (Table S1), the likely deaminase source for ZIKV (acute) and HCV (acute and chronic) MC3 C-transition mutations were identified. Table 3 shows the results for C>U, G>A, A>G and U>C transitions.

Transition mutations, C-to-U and A-to-I(G), targeting MC3 nucleotides off known motifs (Sharma et al 2015, 2016; Rosenberg et al 2011; Bass 2002; Eifler et al 2013) associated with the RNA targeted deaminase editing by APOBEC1, APOBEC3A and ADAR1/2 account for the majority of identified mutations. For ZIKV, HCVacute and HCVchronic the proportion of MC3 transitions at the motifs shown in Table 3 are 76.4%, 60.2% and 63.8% respectively. For the set of NCC motifs not identified with a known APOBEC deaminase they could easily be redundant APOBEC1 and APOBEC3A sites, although further work is required to resolve this. However the great majority of C>U

deamination events at MC3 can be associated with an established motif of either APOBEC1 or APOBEC3A deaminase (or both). Similarly, the vast majority of MC3 A>G (U>C) transitions can also be accounted for in the context of a known W-containing 5′ and 3′ nearest neighbour ADAR deaminase motifs (Bass 2002; Eifler et al 2013).

Given that there are many nonsynonymous transitions at the MC1 and MC2 positions in the HCVc data set, MC1 and MC2 targeted transitions were also grouped by possible RNA APOBEC1, APOBEC3A and ADAR1/2 deaminase motifs. The results are shown in Table 4. As before the majority of MC1 and MC2 transitions can be accounted for by being located at a known APOBEC1 or APOBEC3A motifs, or with known ADAR A-to-I deamination motifs.

By way of comparison, using the mutation data and nucleotide context adjacent to each mutation (Table S1), the number of MC3 C>U/G>A and A>G/U>C transition mutations occurring off motifs known to be associated with APOBEC1, APOBEC3A, and ADAR1/2 were tabulated for the HBV non-overlapping reading frame dataset. The results are shown in Table 5.

Table 4. Number of MC1 and MC2 C>U/G>A and A>G/U>C transition mutations occurring off motifs known to be associated with APOBEC1, APOBEC3A and ADAR1/2 targeting deaminase activity for HCV chronic whole RNA genomes.

Likely Deaminase	C-to-U Motif (+ve RNA Strand)	HCVc	G-to-A Motif (-ve RNA Strand)	HCVc
APOBEC3A/1*	C<u>C</u>A	1	U<u>G</u>G	0
APOBEC3A	U<u>C</u>A	1	U<u>G</u>A	0
	A<u>C</u>A	0	U<u>G</u>U	1
	G<u>C</u>A	0	U<u>G</u>C	2
	C<u>C</u>G	0	C<u>G</u>G	1
	U<u>C</u>G	1	C<u>G</u>A	1
	A<u>C</u>G	2	C<u>G</u>U	3
	G<u>C</u>G	4	C<u>G</u>C	1
APOBEC1	C<u>C</u>U	4	A<u>G</u>G	0
	U<u>C</u>U	1	A<u>G</u>A	0
	A<u>C</u>U	1	A<u>G</u>U	1
	G<u>C</u>U	0	A<u>G</u>C	2
	A-to-G Motif (+ve RNA Strand)		U-to-C Motif (-ve RNA Strand)	
ADAR1/2	W<u>A</u>	7	<u>U</u>W	6
	WS<u>A</u>	7	<u>U</u>SW	3
	<u>A</u>W	2	W<u>U</u>	0
	<u>A</u>SW	0	WS<u>U</u>	0
	Total	31		21

Proportion of MC1/2 transitions off above motifs is (31+ 21)/94 = 55.3%
Edited nucleotide is underlined. *U<u>C</u>A/U<u>G</u>A is a motif targeted by both APOBEC1 and APOBEC3A
[t]W=A/U,S=G/S

Table 5. Number of MC3 C>U/G>A and A>G/U>C transition mutations occurring off tri-nucleotide motifs known to be associated with APOBEC1, APOBEC3A and ADAR1/2 targeting deaminase activity for HBV non-overlapping reading frame dataset.

Likely Deaminase	C-to-U Motif (+ve RNA Strand)	HBVc	G-to-A Motif (-ve RNA Strand)	HBVc
APOBEC3A/1*	CCA	2	UGG	0
APOBEC3A	UCA	1	UGA	0
	ACA	1	UGU	0
	GCA	0	UGC	1
	CCG	1	CGG	0
	UCG	1	CGA	0
	ACG	0	CGU	0
	GCG	0	CGC	0
APOBEC1	CCU	0	AGG	2
	UCU	4	AGA	3
	ACU	0	AGU	0
	GCU	1	AGC	2
	A-to-G Motif (+ve RNA Strand)		U-to-C Motif (-ve RNA Strand)	
ADAR1/2	WA	1	UW	6
	WSA	0	USW	2
	AW	0	WU	0
	ASW	1	WSU	0
	Total	13		14

Proportion of MC1/2 transitions off above motifs is (13+ 14)/40 = 67.5%
Edited nucleotide is underlined
*UCA/UGA is a motif targeted by both APOBEC1 and APOBEC3A
[t]W=A/U,S=G/S

It is found that, although the numbers are small, 67.5% of MC3 transitions can be accounted for by being located at a known APOBEC1 or APOBEC3A motifs, or with known ADAR A-to-I mediated deamination motifs.

To help us to better understand whether potential virus-directed virulence factors are guiding host deamination events to enrich mutations at MC3 C-sites and A-sites a summary of the expected proportions of mutations at MC1, MC2, MC3 based on an idealised process of simple "Random Deamination" producing *only* transition mutations (C>U, G>A, A>G, U>C) with no resulting selection was developed. The expected proportions for transition mutations under a Random Deamination model were tabulated and compared using codon usage data for the open reading frame sequences of the Cambodian 2010 ZIKV strain, with a total of 9,799 possible single nucleotide transitions (Mlakar et al 2016), and the HCV Acute 10,5431 TF1 strain with a total of 9,118 possible transitions. The tabulation and calculations are shown in Table S3 and final results displayed in Figure 2. To simplify the analysis under the random deamination model just defined, it is first noted that in Table 2 nearly

all MC3 transitions are "silent" and have no functional consequence (apart from the small numbers of MC3 transitions at Ile/Met codons and Trp/Stop codons, which are accounted for in the model). It can be seen that many of the non-synonymous transitions at MC1, MC2 sites could potentially be lethal due to an amino acid replacement. However, it has been shown in direct tests for critical levels of lethality by Sanjuan et al (2004) using random site-directed mutagenesis of RNA viruses that are, in principle, similar to the positive strand RNA *Flaviviruses*, that up to 40% of such amino acid changes could be lethal. This is the best comparison available at present (and see the discussion on this point in the Introduction). In our model from Table 2 for ZIKV, the enriched fraction of MC3 synonymous transitions per total transitions is 177/195 or 90.8%; for HCVacute this enriched fraction is 61/83 or 73.5%. If selection is then applied for the expectation of a random deamination transition model, by assuming that a rising fraction (viz. no selection through to 90% selection) of non-synonymous transitions are lethal, then it is possible to estimate the expected enrichment at MC3 as a function of the fraction of non-synonymous transitions (at MC1, MC2) which result in a lethal amino acid replacement. (The calculation steps and tabulation of the results shown are included in Excel Supplementary Data: Table S3 which show the calculations for the selection against MC1/MC2 non-synonymous lethals).

Figure 2 provides a graphical representation of these theoretical results.

Under a Random Deamination model, the expected number of mutations targeting MC3 is around 30%. If you progressively remove a fraction of the non-synonymous totals (presumably as 'lethals'), then the expected enrichment in the number of synonymous mutations targeting MC3 is steadily increased. However, the actual proportions of deamination transitions enriched at MC3 in Table 2 is 90.8% for ZIKV and 73.5% for the HCVa data. To achieve these MC3 enrichment levels 90% of the MC1/MC2 non-synonymous for ZIKV would need to be deleted, and somewhere between 60%-80% for HCVacute.

Whilst a direct comparison with the experimental "40% lethality rule" of Sanjuan et al (2004) is not possible, this simple modelling nevertheless reveals a possible large gap between the expectations under selection and what is observed with respect to MC3 enriched transition deaminations in the acute phase of ZIKV and HCV infections. This opens the door to consider that other mechanisms are influencing the extreme MC3-enriched pattern found in this study.

DISCUSSION

Three main results bear on the origin of mutations in RNA viruses.

First, in the acute phase of infection of both ZIKV, HCV and HBV (non-overlapping regions), the majority numbers of mutations are largely located at the MC3 position within a codon, and the vast majority of these (> 95%) are transitions: C>U, G>A, A>G and U>C. While a similar pattern is observed in the chronic HCV dataset, it is also observed here that there is an increase in nucleotide diversity involving numerous nonsynonymous transversions now located at MC1 and MC2. This pattern is consistent with powerful selection via a fully active adaptive T and B cell immune response expected to be in full swing at this late stage of the chronic infection (Stoddard et al. 2015).

Second, the vast majority of transition mutations, whether at MC3, or at MC1/2 in the HCV chronic data sets, are found to be located at motifs typical of APOBEC1- and APOBEC3A-mediated C-to-U RNA editing; or ADAR1/2-mediated A-to-I RNA editing events at W-containing motifs. This result suggests that the majority of mutations in *Flavivirus* variants and quasi-species population variants can be accounted for by typical RNA

editing events which have been faithfully propagated and expanded in the viral population by the action of a high fidelity viral replicase (RdRP). This further implies that the minority population of rare nonsynonymous transversion mutations evident particularly in the HCV Chronic dataset (Table 2C) may have arisen in frequency under intense immune selection pressure of rare

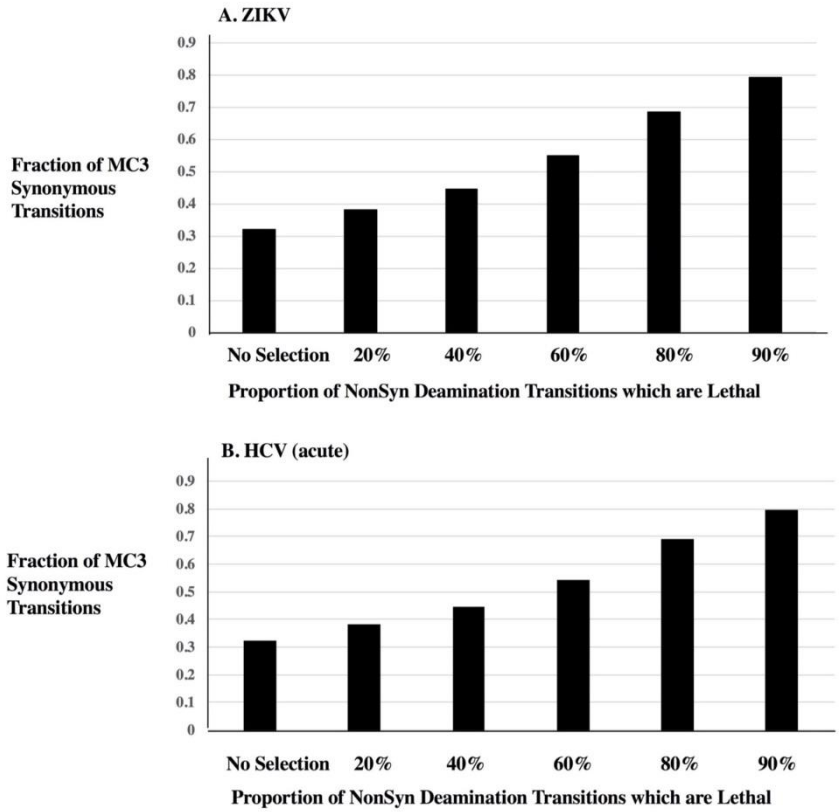

Figure 2. Expected Proportion of MC3 Synonymous Transitions under a Random Deamination Model Producing Only Transition Mutations with various Levels of Selection Against Non-Synonymous Mutations (Table S3)

variants caused by error-prone replicase action since RdRP, although likely to be of relatively high fidelity (Freistadt et al. 2007; Aggarwal et al. 2010), would still lack proof reading functions. Thus with respect to current views on the origin of mutations in RNA viruses, it seems that host deamination events dominate viral mutation patterns, particularly in the early Innate Immunity phase, and that the residue of intrinsic viral mutations are those due to low frequency errors let through by RdRP and thus selectively expanded for viral fitness. An alternative explanation might rely on the theoretical G•U mispair/misincorporation process acting in concert *in vivo* with the host innate deamination response itself causing unbalanced NTP precursor pools as a consequence of the inflammatory insult such as might occur during the infection phase.

Third, it is estimated that there is a possible unexplained gap between the expected proportion of synonymous transition mutations targeting MC3 under a pure Random Deamination model coupled to a purifying selection model where the virus undergoes intense selection for viral survival (Sanjuan et al 2004; Sanjuan 2010), and what is actually observed: the intense selection for virus survival may only partially explain the targeted nature of the results. While 'purifying selection' of the compact integrated information-rich viral

genome that is sensitive to random mutation is important, here it is reasoned that other site-directed mutation mechanisms associated with functions expressed in the *Flaviviruses* ZIKV and HCVacute could be involved.

Although further work is required, the TSM methodology has shown that in the early ZIKV or HCVacute infection phase, many targeted new mutations are the result of deaminase-driven activity occurring in the viral RNA genome. While other targeting mechanisms may also be involved, the result is an effective mutation-driven immune evasion strategy that helps to ensure viral survival.

On this point it is important to consider more deeply the informative site-directed mutagenesis experiments of Sanjuan and associates (Sanjuan et al 2004; Sanjuan 2010; and Pers. Comm. 8 and 22 Aug 2016, permission granted to quote): Thus, " the sites for mutations along the viral genomes were chosen at random, and while it is true that introduced mutations differ in some way from the actual spectrum of spontaneous mutations produced by a given virus (as it is known these spontaneous mutations appear to be biased such that transitions outnumber transversions) the sets of introduced site-directed mutations were completely random in contrast to the known natural biases spectrum of mutations seen in RNA viruses." And according to Rafel Sanjuan "...the 40% rule is probably an upper limit. After analysing a few different viruses, they found that the fraction of lethal mutations ranged 20-40%, as detailed in Sanjuan (2010) which is still a very high fraction compared to other biological systems."

The main conclusion is that site-directed host-mediated deaminase editing plays a key role in the viral lifecycle. In this context, it appears that most of the viral mutations giving rise to new variants are due to a virus-host contest for survival that involves targeted deaminase activity that is then acted upon by intense selection to produce a quasi-species population with a predominant number of synonymous mutations targeting MC3 sites. Synonymous changes at 3rd positions in codons (MC3) were for a long time considered "silent" and without functional consequence. However, over the past two decades a body of literature now clearly shows that synonymous changes at MC3 can have significant functional consequences on the protein products of such mutated genes (Drummond & Wilke 2008; Goodman et al 2013; Buhr et al 2016; Pirakitikulr et al 2016). These data are of particular relevance to the MC3 C-to-U and A-to-I transitions in the information dense polyprotein coding sequences of viruses such as the *Flaviviruses* considered here. Thus, co-translational protein mis-folding caused by altered rates of protein extension from the ribosome (due to rarely used synonymous codons) can have significant impact resulting in proteins with altered or different functions.

These new findings might be taken into account in developing new therapeutic intervention strategies (e.g. in druggable pocket modelling and prediction, Cox et al 2015), as well as in the identification and development of more targeted oligonucleotide "deaminase blockers" in association with the emerging crystal structures of the anti-viral genome deaminases. Thus, for the RNA and RNA/RT-dependent viruses (HCV, HBV, HIV-1) a new focus might be on the role of ADAR and APOBEC3 RNA editing deaminases (ADAR1, ADAR2, APOBEC1, APOBEC3A) – quite apart from the well-studied APOBEC3G DNA deaminase - on restriction of viral replication and how there may be a subtle, hitherto unexpected, viral evasion strategy based on MC3-targeted deamination which may well be of vital functional consequence, and thus survival value, for the virus. That is, the implication for intervention and drug discovery is that conserved regions of amino acid sequence within variant viral strains should now be considered as potential druggable targets. i.e. hypermutable MC3-deaminase targeted codons in the viral genome.

Finally, the assumption that 3rd position (MC3) synonymous changes are "silent" has also been a

central canon in evolutionary genetic thinking for over three decades. Kimura's *Neutral Theory of Molecular Evolution* (Kimura 1983) has been built on this assumption: Changes in the third position which are silent cannot be selected for a new or modified protein function. This allowed the molecular evolution field to rely on some important assumptions: a) The Synonymous to non-synonymous ratio in mutation or polymorphism analyses to be used as an indicator of so called "positive selection", and, b) The non-selected "silent" mutation rates can be used in molecular evolution studies as "molecular clocks" in establishing the time of divergence from Last Common Ancestor (LCA) in, for example, protein evolutionary trees.

Further, it is argued that both of these assumptions now need to be qualified by the work reported here, and that of Buhr et al (2016) which is the result of two decades of careful work by the Komar group. Other important papers to read in this context are Drummond & Wilke 2008; Goodman et al 2013. In essence, because the back-up pools of nucleotide and other precursors for a given codon (tRNAs, amino-acyl tRNA synthetases etc.) will vary between tissues (and organisms) a silent base switch will have no functional consequence on the structure of the folded protein provided that these precursors pool sizes are sufficient to prevent co-translational pausing or delays during exit of the protein from the Ribosome and thus folding kinetics. If these are out of sync Buhr et al (2016) show that mis-folding resulting in new protein function can result.

In addition it has also been reported recently that the HCV RNA genome contains a number of well-conserved RNA structures within the protein-coding regions, and that genetic manipulation of these regions alters the ability of HCV to replicate, and hence infect (Pirakitikulr et al 2016). Targeting subtle variations in RNA sequence in these conserved RNA secondary structures, at the hypermutable MC3 positions, may be good proxies for "druggable pockets" within the RNA itself rather than at the protein level.

Thus it is now becoming more evident from the work of others that viral RNA genomes contain a number of additional layers of information that impact viral potency and function. Therefore, traditional silent 3rd position changes within the body of an exon can result in significant functional consequences, and further investigation is encouraged to better understand their impact on viral replication and infectivity.

ACKNOWLEDGEMENT

We thank Professor Stephen Locarnini and Dr Nathan Hall for insightful comments and feedback on the initial draft of the manuscript.

CONFLICT OF INTEREST

Authors declares there is no conflict of interest

REFERENCES

Aaskov, J., Buzacott, K., Thu, H.M., Lowry, K., and Holmes, EC. (2006). Long- term transmission of defective RNA viruses in humans and Aedes mosquitoes. Science *311*, 236-238. DOI: 10.1126/science.1115030

Aggarwal, S., Bradel-Tretheway, B., Takimoto, T., Dewhurst, S., and Kim, B. (2010). Biochemical characterization of enzyme fidelity of influenza A virus RNA polymerase complex. PLoS ONE ***5****(4)*, e10372. doi:10.1371/journal.pone.0010372

Andino, R., and Domingo, E. (2015). Viral quasispecies. Virology *479-480*, 46-51. DOI: 10.1016/j.virol.2015.03.022

Armitage, A.E., Deforche, K., Welch, J.J., van Laethem, K., Camacho, R., Rambaut, A., Astrid, K.N., and Iversen, A.K.N. (2014). Possible footprints of APOBEC3F and/or other APOBEC3 Deaminases, but not APOBEC3G, on HIV-1 from patients with acute/early and chronic infections. J. Virol. *88*, 12882-12894. DOI: 10.1128/JVI.01460-14

Barrett, B.S., Guo, K., Harper, M.S., Li, S.X., Heilman K.J., Davidson, N.O., Santiago, M.L. (2014). Reassessment of murine APOBEC1 as a retrovirus restriction factor *in vivo* Virology *468-470,* 601-608. DOI: 10.1016/j.virol.2014.09.006

Bass, B.L. (2002). RNA editing by adenosine deaminases that act on RNA. Ann Rev Biochem *71*, 817-846. DOI: 10.1146/annurev.biochem.71.110601.135501

Bohn, M-F., Shandilya, S.M.D., Albin, J.S., Kouno, T., Anderson, B.D., McDougle, R.M., Carpenter, M.A., Rathore, A., Evans, L., Davis, A.N., Zhang, J., Lu, Y., Somasundaran, M., Matsuo, H., Harris, R.S., and Schiffer, C.A. (2013). Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically activeand HIV-1 Vif-binding domain. Structure *21*, 1042-1050. DOI: 10.1016/j.str. 2013.04.010

Bourara, K., Litvak, S., and Aray A. (2000). Generation of G-to-A and C-to-U changes in HIV-1 transcripts by RNA editing. Science *289*, 1564-1566. PMID: 10968794

Bridges, B.A. (2005). Error-prone DNA repair and translesion DNA synthesis. II: The inducible SOS hypothesis. DNA Repair *4*,725-726, 739. DOI: 10.1016/ j.dnarep.2004.12.009

Buhr, F., Jha, S., Thommen, M., Mittlestael, J., Kutz, F., Schwalbe, H., Rodina, M.V., and Komar, A.A. (2016). Synonymous codons direct cotranslational folding towards different protein conformations. Mol. Cell 61, 341-351. http://dx.doi.org/10.1016/j.molcel.2016.01.008

Carpenter, M.A., Li, M., Rathore, A., Lackey, L., Law, E.K., Land, A.M., Leonard, B., Shandilya, S.M.D., Bohn, M-F., Schiffer, C.A., Brown, W.L., and Harris, R.S. (2012). Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. J. Biol Chem *287*, 34801-34808. DOI: 10.1074/jbc.M112.385161

Casey, J.L., and Gerin, J.L. (1995). Hepatitis D virus RNA editing: Specific modification of adenosine in the antigenomic RNA. Virology *69*, 7593-7600. PMCID: PMC189698

Chen, Y-H., Du, W.L., Hagemeijer, M.C., Takvorian, P.M., Pau, C., Cali, A., Branter, C.A., Stempinski, E.S., Connelly, P.S., Ma, H.C., Jiang, P., Wimmer, E., Altan-Bonnet, G., Altan-Bonnet, N. (2015). Phosphatidylserine vesicles enable efficient *en bloc* transmission of enteroviruses. Cell 160, 619-630. http://dx.doi.org/10.1016/j.cell.2015.01.032

Combe, M., Garijo, R., Geller, R., Cuevas, J.M., and Sanjuan, R. (2015). Single-cell analysis of RNA virus infection identifies multiple genetically diverse viral genomes within single infectious units. Cell Host & Microbe *18*, 424-432. DOI: 10.1016/j.chom.2015.09.009

Conticello, S.G. (2008) The AID/APOBEC family of nucleic acid mutators. Genome Biology *9*, 229 doi:10.1186/gb-2008-9-6-229.

Cox, B.D., Stanton, R.A., and Schinaz.i R.F. (2015). Predicting Zika virus structural biology: challenges and opportunities for intervention. Antiviral Chem.Chemotherapy. *24*, 118-126. DOI: 10.1177/2040206616653873

Drummond, D.A., and Wilke, C.O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding sequence evolution. Cell *134*, 341-352. DOI: 10.1016/j.cell.2008.05.042

Eifler, T., Pokharel, S., and Beal, P.A. (2013). RNA-Seq analysis identifies a novel set of editing substrates for human ADAR2 present in *Saccharomyces cerevisiae*. Biochemistry *52*, 7857-7869. DOI: 10.1021/bi4006539

Eigen, M., and Schuster, P. (1979). The Hypercycle: A Principle of Natural Self-Organization. Springer, Berlin.

Faye, O., Freire, C.C.M., Lamarino, A., Faye, O., de Oliveira, J.V.C., Diallo, M., Zanotto, P.M., and Sall AA. (2014). Molecular evolution of Zika virus during its emergence in the 20th century. PLoS Negl Trop Dis *8(1)*, e2636. doi: 10.1371/journal.pntd.0002636

Franklin, A., Milburn, P.J., Blanden, R.V., and Steele, E.J. (2004). Human DNA polymerase- η an A-T mutator in somatic hypermutation of rearranged immunoglobulin genes, is a reverse transcriptase. Immunol. Cell Biol. *82*, 219 - 225. DOI: 10.1046/j.0818- 9641.2004.01221.x

Freistadt, M.S., Vaccaro, J.A., and Eberle, K.E. (2007). Biochemical characterization of the fidelity of poliovirus RNA-dependent RNA polymerase. Virology J. *4*,44 doi:10.1186/1743-422X-4-44

Gonzalez, M.C., Suspène, R., Henry, M., Guétard, D., Wain-Hobson, S., and Vartanian, J.P. (2009). Human APOBEC1 cytidine deaminase edits HBV DNA. Retrovirology *6*,96. doi: 10.1186/1742-4690-6-96.

Goodman, M.F. (2002). Error-prone repair DNA polymerases in prokaryotes and eukaryotes. Annu. Rev. Biochem. *71*, 17-50.DOI:10.1146/annurev.biochem.71.083101.124707

Goodman, B.G., Church, G.M., and Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. Science *342*, 475-479. DOI: 10.1126/science.1241934

Hajjar, A.M., and Linial, M.L. (1995). Modification of retroviral RNA by double-stranded RNA adenosine deaminase. Virology *69*, 5878-5882. PMCID: PMC189466

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003). DNA deamination mediates innate immunity to retroviral infection. Cell *113*, 803-809. PMID: 12809610

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and Vandepol, S. (1982). Rapid evolution of RNA genomes. Science *215*, 1577-I 585. PMID: 7041255

Keele, B.F., Giorgi, E.E., Salazar-Gonzalez, J.F., Decker, J.M., Pham, K.T., Salazar, M.G., Sun, C., Grayson, T., Wang, S., Li, H., et al. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. USA *105*, 7552-7557 DOI: 10.1073/pnas.0802203105

Kim, T., Mudry, R.A., Rexrode, C.A., and Pathak, V.K. (1996). Retroviral mutation rates and A-to-G hypermutations during different stages of retroviral replication. Virology *70*, 7594-7602. PMCID: PMC190828

Kimura, M. (1983). The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Kock, J., and Blum, H.E. (2008). Hypermutation of hepatitis B virus genomes by APOBEC3G, APOBEC3C and

APOBEC3H. J Gen. Virol. *89*, 1184-1191. DOI: 10.1099/vir.0.83507-0

Kuttan, A., and Bass, B.L. (2012). Mechanistic insights into editing-site specificity of ADARs. Proc. Natl. Acad. Sci. USA. *109*, E3295–E3304. DOI: 10.1073/pnas.1212548109

Lauring, A.S., Frydman, J., and Andino, R. (2013). The role of mutational robustness in RNA virus evolution. Nat. Rev. Microbiol *11*, 327–336. DOI: 10.1038/nrmicro3003

Lecossier, D., Bouchonnet, F., Clavel, F., and Hance, A.J. (2003). Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science *300*, 1112. DOI: 10.1126/science.1083338

Li, H., Stoddard, M.B., Wang, S., Blair, L.M., Giorgi, E.E., Parrish, E.H., Learn, G.H., Hraber, P., Goepfert, P.A., Saag, M.S., et al. (2012). Elucidation of hepatitis C virus transmission and early diversification by single genome sequencing. PLoS Pathog *8(8)*, e1002880. doi: 10.1371/journal.ppat.1002880

Lindley, R.A. (2013). The importance of codon context for understanding the Ig-like somatic hypermutation strand-biased patterns in TP53 mutations in breast cancer. Cancer Genetics *5*, 2619-2640. DOI: 10.1016/j.cancergen. 2013.05.016

Lindley, R.A., and Hall, N.E 2018. (2018) APOBEC and ADAR deaminases may cause many single nucleotide polymorphisms curated in the OMIM database. Accepted for Publication, In Press. Mutation Research March 2018

Lindley, R.A., Humbert, P., Larmer, C., Akmeemana, E.H., and Pendlebury, C.R.R. (2016). Association between targeted somatic mutation (TSM) signatures and HGS-OvCa progression. Cancer Med. *5*, 2629-2640. DOI: 10.1002/cam4.825

Logue, E.C., Bloch, N., Dhuey, E., Zhang, R., Cao, P., Herate, C., Chauveau, L, Hubbard, S.R., and Landau, N.R. (2014). A DNA sequence recognition loop on APOBEC3A controls substrate specificity. PLoS ONE *9(5)*, e97062. doi:10.1371/journal.pone.0097062

Martınez, I., and Melero, J.A. (2002). A model for the generation of multiple A to G transitions in the human respiratory syncytial virus genome: predicted RNA secondary structures as substrates for adenosine deaminases that act on RNA. J Gen. Virol. *83*, 1445–1455. DOI: 10.1099/0022-1317-83-6-1445

Mlakar, J., Korva, M., Tul, N., Popovic, M., Poljsak-Prijatej,M., Mraz J, Kolenc, M., Resman Rus, K., Vesnaver Vipotnik, T., Fabjan Vodušek, V., Vizjak, A., Pižem J., Petrovec, M., and Županc TA.(2016). Zika virus associated with microcephaly. New Engl. J. Med. 10,1-8. DOI: 10.1056/NEJMoa1600651

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 424, 99-103. DOI: 10.1038/nature01709

Monajemi, M., Woodworth, C.F., Benkaroun, J., Grant, M., and Larijani, M. (2012). Emerging complexities of APOBEC3G action on immunity and viral fitness during HIV infection and treatment Retrovirology *9*, 35 http://www.retrovirology.com/content/9/1/35

Murakami, E., Feng, J.Y., Lee, H., Hanes, J., Johnson, K.A., and Anderson, K.S. (2003). Characterization of novel reverse transcriptase and other RNA-associated catalytic activities by human DNA polymerase– γ. J. Biol. Chem. *278*, 364013–36409. DOI: 10.1074/jbc.M306236200

Murphy, D.G., Dimock, K., and Kang, C.Y. (1991). Numerous transitions in human parainfluenza virus 3 RNA recovered from persistently infected cells. Virology *181*, 760–776. PMID: 1849685

Naraiza,I., Linfesty, D.C., Greener, B.N., Hakata, Y., Pintel, D.J., Logue, E., Landau, N.R., and Weitzman, M.D. (2009). Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase. PLoS Pathog *5(5)*, e1000439. doi:10.1371/journal.ppat.1000439

Pirakitikulr, N., Kohlway, A., Lindenbach, B.D., and Pyle, A.M. (2016). The coding region of the HCV genome contains a network of regulatory RNA structures. Mol. Cell *61*, 1-10. http://dx.doi.org/10.1016/j.molcel.2016.01.024

Pourkarim, M.R., Verbeeck, J., Rahman, M., Amini-Bavil-Olyaee, S., Forier, A-M., Lemey, P., Maes, P., and Ranst, M.V. (2009). Phylogenetic analysis of hepatitis B virus full-length genomes reveals evidence for a large nosocomial outbreak in Belgium. J Clin Virology *46*, 61-68. doi: 10.1016/j.jcv.2009.06.015. Epub 2009 Jul 16

Powdrill, M.H., Tchesnokova, E.P., Kozak, R.A., Russell, R.S., Martin, R., Svarovskaiad, E.S., Mo, H., Kouyos, R.D., and Götte, M. (2011). Contribution of a mutational bias in hepatitis C virus replication to the genetic barrier in the development of drug resistance. Proc. Natl. Acad. Sci. USA *108*, 20509-20513. DOI: 10.1073/pnas.1105797108

Radman, M. (2005) SOS replication: a distinct DNA replication mechanism which is induced by DNA-damaging treatments? 1970. DNA Repair *4*, 732-738. PMID: 15981323

Refsland, E.W., and Harris, R.S. (2013). The APOBEC3 family of retroelement restriction factors. Curr. Top. Microbiol. Immunol. *371*, 1-27. DOI: 10.1007/978-3-642-37765- 5_1

Ribeiro, R.M., Li, H., Wang, S., Stoddard, M.B., Learn, G.H., Korber, B.T., Bhattacharya, T., Guedj, J., Parrish, E.H., Hahn, B.H., Shaw, G.M., and Perelson, A.S. (2012). Quantifying the diversification of hepatitis C virus (HCV) during primary infection: estimates of the *in vivo* mutation rate. PLoS Pathog. *8(8)*, e1002881. doi: 10.1371/journal.ppat.1002881

Rosenberg, B.R., Hamilton, C.E., Mwangi, M.M., Dewell, S., and Papavasiliou, F.N. (2011).

Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA editing targets in transcript 3' UTRs. Nat. Struct. Mol. Biol. *18*, 230-236. DOI: 10.1038/nsmb.1975

Samuel, C.E. (2011). Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. Virology *411*, 180–193. DOI: 10.1016/j.virol.2010.12.004

Sanjuan, R., Nebot, M.R., Chirico, N., Mansky, L.M., and Belshaw, R. (2010). Viral mutation rates. Virology *84*, 9733-9748. DOI: 10.1128/JVI.00694-10

Sanjuan, R. (2010). Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. Phil. Trans. Roy. Soc. B. *365*, 1975–1982. DOI: 10.1098/rstb.2010.0063

Sanjuan, R. (2012). From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. PLoS Pathog. *8(5)*, e1002685. doi: 10.1371/journal.ppat.1002685

Sanjuan, R., and Domingo-Calap, P. (2016). Mechanisms of viral mutation Cell. Mol. Life Sci. *73*, 4433 – 4448. DOI: 10.1007/s00018-016-2299-6

Sanjuan, R., Moya, A., and Elena, S.F. (2004). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc. Natl. Acad. Sci. USA. *101*, 8396 – 8401. DOI: 10.1073/pnas.0400146101

Sharma, S., Patnaik, S.K., Taggart, R.T., Kannisto, E.D., Enriquez, S.M., Gollnick, P., and Baysal, B.E. (2015). APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat. Commun. *6*. 6881. doi: 10.1038/ncomms7881

Sharma, S., Santosh, K.. Patnaik, S.K., Kemera, Z., and Baysal BE. (2016). Transient overexpression of exogenous APOBEC3A causes C-to-U RNA editing of thousands of genes. RNA Biology *5*, 1-8 doi: 10.1080/15476286.2016.1184387

Shi, K., Carpenter, M.A., Kurahashi, K., Harris, R.S., and Aihara, H. (2015). Crystal structure of the DNA deaminase APOBEC3B catalytic domain. J Biol. Chem. *290*, 28120 - 28130. DOI: 10.1074/jbc.M115.679951

Shin, E-C., Park, S-H., Demino, M., Nascimbeni, M., Mihalik, K., Major, M., Veerapu, N.S., Heller, T., Feinstone, S.M., Rice, C.M., and Rehermann, B. (2011). Delayed induction, not impaired recruitment, of specific CD8+ T cells causes the late onset of acute hepatitis C. Gastroenterology *141*, 686-695. DOI: 10.1053/j.gastro.2011.05.006

Steele, E.J., and Lindley RA. (2017). ADAR deaminase A-to-I editing of DNA and RNA moieties of RNA:DNA hybrids has implications for the mechanism of Ig somatic hypermutation. DNA Repair 55, 1-6. DOI: 10.1016/j.dnarep.2017.04.004

Stoddard, M.B., Li, H., Wang, S., Saeed, M., Andrus, L., Ding, W., Jiang, X., Learn, G.H., von Schaewen, M., Wen, J., Goepfert, P.A., Hahn, B.H., Ploss, A., Rice, C.M., and Shaw G.M. (2015). Identification, molecular cloning, and analysis of full-length hepatitis C virus transmitted/founder genotypes 1, 3, and 4. mBio *6(2)*, e02518-14. doi: 10.1128/mBio.02518-14

Sung, P.S., Racanelli, V., and Shin, E-C. (2014). CD8+ T-cell responses in acute hepatitis C virus infection. Frontiers in Immunology *5*, Article 266. doi: 10.3389/fimmu.2014.00266

Taylor, D.R., Puig, M., Darnell, M.E.R., Mihalik, K., and Feinstone, S.M. (2005). New antiviral pathway that mediates hepatitis C virus replicon interferon sensitivity through ADAR1.J. Virol. *79*, 6291–6298. DOI: 10.1128/JVI.79.10.6291-6298.2005

Thimme, R., Binder, M., and Bartenschlager, R. (2012). Failure of innate and adaptive immune responses in controlling hepatitis C virus infection. FEMS Microbiol. Rev. *36*, 663–683. DOI: 10.1111/j.1574-6976.2011.00319.x

Tippin, B., Pham, P., and Goodman, M.F. (2004).Error-prone replication for better or worse. Trends Microbiol. *12*, 288-295. DOI: 10.1016/j.tim.2004.04.004

Vartanian, J-P., Henry, M., Marchio, A., Suspene, R., Aynaud, M-M., Guétard D., Cervantes- Gonzalez, M., Battiston, C., Mazzaferro, V., Pineau, P., Dejean, A., and Wain- Hobson S. (2010). Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. PLoS Pathog. *6(5)*, e1000928. doi: 10.1371/journal.ppat.1000928

te Velthuis, A.J.W. (2014). Common and unique features of viral RNA-dependent polymerases. Cell. Mol. Life Sci. *71*,4403–4420. DOI: 10.1007/s00018-014-1695-z

Xu, R., Zhang, X., Zhang, W., Fang, Y., Zheng, S., and Yu X-F. (2007). Association of human APOBEC3 cytidine deaminases with the generation of hepatitis virus Bx antigen mutants and hepatocellular carcinoma. Hepatology *46*, 1810–1820. DOI: 10.1002/hep.21893

Zahn, R.C., Schelp, I., Utermohlen, O., von Laer, D. (2007). A-to-G hypermutation in the genome of lymphocytic choriomeningitis virus. J. Virol. *81*, 457–464. DOI: 10.1128/JVI.00067-06

Zeng, X., Winter, D.B., Kasmer, C., Kraemer, K.H., Lehmann, A.R., and Gearhart, P.J. (2001). DNA polymerase-η as an A-T mutator in somatic hypermutation of immunoglobulin variable genes, Nat. Immunol. *2*, 537 - 541. DOI: 10.1038/88740

Zheng, Y.C., Lorenzo, C., and Beal, P.A. (2017) DNA Editing in DNA/RNA hybrids by adenosine deaminases that act on RNA, Nucleic Acids Res. *45*, 3369–3377. DOI: 10.1093/nar/gkx050