C-145

Analysis of HIV-1 G→A Hypermutation and it's Relationship with APOBEC3G and *Vif* Genetic Variation *In Vivo*

[1]Craig Pace, [1]David Nolan, [1]Jean Keller, [1]Silvana Gaudieri, [1]Filipa Carvalho, [1]Corey Moore, [1]Elizabeth Freitas, [1]Karen Sutton and [1,2]Simon Mallal

[1]Centre for Clinical Immunology and Biomedical Statistics (CCIBS), Royal Perth Hospital and Murdoch University, Perth, Western Australia

[2]Department of Clinical Immunology and Biochemical Genetics, Royal Perth Hospital, Western Australia.

Contact - Professor Simon Mallal, Executive Director
Centre for Clinical Immunology & Biomedical Statistics
Level 2, R Block, Royal Perth Hospital
Wellington St, WA, 6000. AUSTRALIA
S.Mallal@murdoch.edu.au
Ph: (61) 8 9224 2899, Fax: (61) 8 9244 2920
Website - http://www.ccibs.org

Western Australian HIV Cohort Study

ABSTRACT

Background: Editing of HIV-1 DNA and RNA by APOBEC family cytidine deaminases to produce hypermutated sequences has emerged as a potent host antiviral response. This effect is modulated by the presence of *Vif* that opposes APOBEC activity. Here we examine the clinical relevance and extent of hypermutation within proviral DNA, the sequence context of G→A hypermutation, and the contributions of genetic variation in host APOBEC3G and HIV-1 *Vif*.

Methods: Full-length HIV-1 proviral DNA sequences from 185 antiretroviral-naive patients were examined, including 18,136 G→A substitutions. Hypermutation was classified by 3 criteria and validated by analyses of the population distribution of G→A substitutions. In individuals with hypermutation, APOBEC3G exonic and putative 5' regulatory region sequences were compared with control data (n=200). Analyses of *Vif* amino acid sequences were also undertaken, and pre-treatment viral load (VL) was assessed in group comparisons.

Results: All 3 criteria for hypermutation were satisfied in sequences obtained from 8 (4.3%) individuals, with 1 criterion met by 20 (10.8%). VL (log_{10}copies/mL) was significantly lower in the hypermutated group using the broader definition (4.25 v 4.89, P=0.01) but not the restrictive definition (4.56 v 4.83, P=0.36). Compared with background rates of 6.1±2.5% G→A substitution, hypermutated sequences demonstrated 20.3±6.2% G→A substitutions by a restrictive definition (broad = 16.0±5.5%) (both P<0.001). Hypermutated G→A substitutions demonstrated a preference for G at the +1 position (45% v 23%) and enrichment of T at position -1 (34% v 23%). (P<0.001) consistent with an APOBEC3G plus strand DNA target motif TGC (corresponding to CCA on the minus strand DNA). Expected strong selection against CpG dinucleotides was also universally noted. APOBEC3G sequencing of the 8 subjects with definite hypermutation revealed no novel variants. Analysis of *Vif* sequence however, revealed that 100% of the hypermutated group had defective *Vif*, defined as absence of appropriate start and stop codons and/or presence of in-frame stop codons.

Conclusions: This *in vivo* study demonstrates that hypermutation is not infrequent at a population level and is associated with lower pre-treatment viral load, in keeping with a host antiviral effect potentially mediated by APOBEC3G.

INTRODUCTION

APOBEC3G (3G) and APOBEC3F (3F) are potent inhibitors of HIV-1 infectivity by deaminating an inordinate number of cytidines in the intermediary DNA strand (minus) following reverse transcription, resulting in excessive guanidine to adenine substitutions in the proviral DNA (hypermutation). Hypermutation has been considered to only occur in the absence of *vif*, the viral infectivity protein that complexes with 3G and 3F, promoting their ubiquitination and proteasomal degradation. However recent *in vitro* data indicate 3G and 3F are partially resistant to *vif* [1,2,3] and this partial resistance to *vif* may explain why G/A substitutions are the most common substitution observed in lentiviral DNA. APOBEC3G and -3F preferentially deaminate cytidines in the CCA and TC context, respectively [4,5] and we used this sequence specificity to investigate the contribution of each to HIV-1 G→A substitutions retrospectively in near full-length HIV-1 proviral DNA sequences from 185 HIV+, antiretroviral naïve participants in the Western Australian HIV-1 cohort (6,416 ± 1,998 bases per patient) and their subsequent influence on viral load.

METHODS

• Retrospective analysis of 18, 674 G→A substitutions (relative to population consensus) in near full-length HIV-1 proviral sequences (6,416 ± 1,998 bases per patient) derived from 185 patients. Sequence context of substitutions was assessed by comparing frequencies of specific nucleotides at nucleotide positions -2 to +2 relative to deamination sites in hypermutated patients versus non-hypermutated patients.

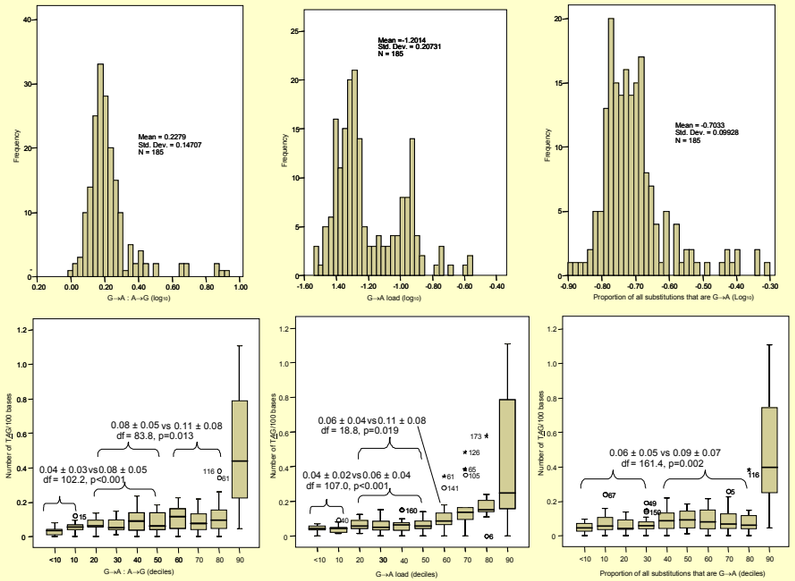
• Proviral DNA was analysed according to deciles of G→A substitution characteristics by ANOVA with multiple comparisons and Bonferroni corrections.

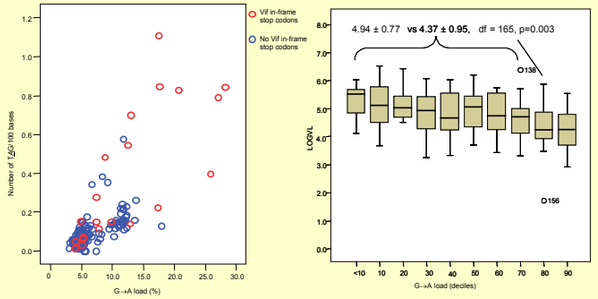
• Contribution of APOBEC3G allelic variation to hypermutation was assessed by comparison of the allele frequencies of patients with >42% frequency of G →A substitutions in the GG context (n=8) to a pooled DNA sample (n=200), estimated by chromatograph relative peak heights, by Fishers exact test.

• The contribution of functional versus non-functional *vif* to hypermutation was determined by comparing the presence of in-frame stop codons to G→A load and number of G→A substitutions in the TAG context.

G→A Mutational Load Characteristics

G→A load Decile	G→A load	-1 position A%	-1 position C%	-1 position G%	-1 position T%	+1 position A%	+1 position C%	+1 position G%	+1 position T%	%TAG of G→A substitutions	TAG/100 bases	LOGVL
<0	<3.9%	56.5	4.9	19.8	18.8	39.0	21.0	20.3	19.6	5.6	0.0394	5.291
1	4.2%	58.1	6.9	14.1	21.0	37.5	22.7	18.7	21.1	4.8	0.0388	5.112
2	4.6%	52.6	7.7	16.3	23.4	35.8	21.5	22.1	20.7	6.9	0.0645	5.145
3	4.9%	57.0	7.2	14.7	21.1	40.4	20.6	20.2	18.9	6.0	0.0610	4.835
4	5.2%	51.4	8.0	16.7	23.9	36.1	23.4	21.0	19.4	5.9	0.0633	4.719
5	5.6%	52.6	7.6	15.7	24.2	35.2	21.8	20.3	22.8	5.1	0.0614	4.972
6	7.9%	49.5	7.4	17.6	25.5	37.7	17.6	28.2	16.5	8.0	0.1139	4.795
7	10.6%	52.2	7.8	16.5	23.5	35.9	17.5	29.7	16.9	8.7	0.1675	4.687
8	12.0%	51.7	8.1	17.4	22.8	35.1	20.7	26.6	17.6	6.7	0.1778	4.385
9	≥12.0%	51.1	6.2	15.5	27.2	41.2	13.8	32.6	12.3	11.2	0.4264	4.244
Total	-	53.3	7.2	16.4	23.1	37.4	20.1	24.0	18.6	6.9	0.1209	4.818
f value	-	2.755	1.457	1.291	2.571	1.286	3.918	5.527	4.276	3.052	16.319	3.512
p value	-	0.005	0.167	0.245	0.010	0.247	<0.001	<0.001	<0.001	0.002	<0.001	0.001





APOBEC3G Allele Frequencies

rs#	Control (n=200) Position	Control (n=200) Allele 1	Control (n=200) Allele 2	Do et al, 2005 Allele 2	Hypermutated (n=8) Allele 1	Hypermutated (n=8) Allele 2	A1A1	A1A2	A2A2
7291971	-962	0.60	0.40	0.40	0.44	0.56	2	3	3
6519166	-959	0.70	0.30	0.31	0.56	0.44	2	5	1
57507473	-90	0.50	0.50	0.28	0.44	0.56	1	5	2
12160242	2664	>0.90	<0.10	0.09	0.94	0.06	7	1	0
9607609	2809*	0.80	0.20	-	0.38	0.63	2	2	4
5757465	F119F	0.60	0.40	0.51	0.75	0.25	4	4	0
12158985	8169	>0.90	<0.10	0.07	0.88	0.12	6	2	0
2294366	8170	0.70	0.30	0.34	0.56	0.44	2	5	1
2294367	8391	0.60	0.40	0.42	0.44	0.56	2	3	3
-	9032	>0.90	<0.10	-	0.88	0.12	6	2	0
3091374	9062	0.50	0.50	-	0.44	0.56	2	3	3
5757467	9076*	0.70	0.30	-	0.31	0.69	0	5	3
5757468	9077	0.70	0.30	-	0.56	0.44	2	5	1
-	9078	>0.90	<0.10	-	0.88	0.12	6	2	0
-	Q275E	>0.90	<0.10	-	0.88	0.012	6	2	0
5757471	11412	>0.90	<0.10	-	0.94	0.06	7	1	0
17537574	11444	>0.90	<0.10	0.07	0.88	0.12	6	2	0
17537581	11626	>0.90	<0.10	0.12	0.88	0.12	6	2	0
-	+22*	0.50	0.50	-	0.92	0.06	5	1	0
-	+28	0.70	0.30	-	1.00	0.0	8	0	0
5757472	+219	0.50	0.50	-	0.44	0.56	1	5	2
3891126	+430	>0.90	<0.10	-	0.88	0.12	6	2	0

Results

• Overall, proviral sequences had 1.83 ± 1.05-fold more G→A mutations than A→G mutations (range 1.0 - 8.7), accounted for 20.4 ± 5.8% (range 12.9% - 48.9%) of all mutations, and 7.2 ± 4.3% (range 2.9% - 28.2%) of consensus G's were mutated to A.

• The frequency of consensus Gs mutated to A appeared bimodally distributed at approximately 4% and 12%.

• Each mutational characteristic was significantly positively correlated to the frequency of G→A mutations specifically in the TGG context (r2 0.737-0.830, p<0.001), GG context (r2 0.436-0.512, all p<0.001), TG context (r2 0.221-0.382, all p<0.003), and negatively correlated to viral load (r2 -0.178 - -0.315, p<0.015).

• A significant increase in the frequency of G→A mutations in the GG and TG contexts occurred in the 60th percentile based on consensus Gs mutated to A values (GG 29.2% vs 20.4%, p=0.028; TG 25.5% vs 22.1%, p=0.024).

• A significant increase in the number of G→A mutations the TGG context was evident at the 20th percentile (0.06 ± 0.04 vs 0.04 ± 0.02 100^{-1} bp), the 60th percentile (0.11 ± 0.08 vs 0.06 ± 0.04 100^{-1} bp, p=0.019) and again at the 90th percentile (0.43 ± 0.34 vs 0.15 ± 0.11 100^{-1} bp, p=0.003) based on number of consensus Gs mutated to A values or G→A: A→G values (20th 0.08 ± 0.05 vs 0.04 ± 0.03 100^{-1} bp, p<0.001; 60th 0.11 ± 0.08 vs 0.08 ± 0.05 100^{-1} bp, p=0.013; 90th 0.48 ± 0.31 vs 0.11 ± 0.08 100^{-1} bp, p<0.001).

• A significant increase in the number of G→A mutations the TGG context was evident at the 40th percentile (0.09 ± 0.07 vs 0.06 ± 0.05 100^{-1} bp, p=0.002) and 90th percentile (0.48 ± 0.30 vs 0.09 ± 0.07 100^{-1} bp, p<0.001) based on proportion of G→A mutations relative to all other mutations.

• A significant reduction in viral load was evident at the 80th percentile based on G→A load (log_{10} 4.37 vs 4.94, p=0.003).

• In multivariate linear regression analysis of viral load including protective host factors HLA-B27, -B57, -B58 and CCR5Δ32, G→A load (β -0.337, P<0.001), CCR5Δ32 (β -0.207, P=0.003) and HLA-B57 (β -0.153, P=0.027) remained highly significant.

• Significant differences in the allele frequencies of 3 non-coding APOBEC3G SNPs were observed in 8 patients harbouring hypermutated proviral sequence compared to a pooled DNA control sample (n=200), although the functional significance is unclear.

• 21 (12.7%) of participants had stop codons in the vif ORF. These participants had significantly higher G→A substitution characteristics (all p<0.002), frequency of G→A substitutions in the TAG (0.38 ± 0.34 vs 0.09 ± 0.08 100^{-1} bp, p=0.001) and GG (33.0 ± 14.0% vs 22.8 ± 8.1%, p=0.003) contexts and significantly lower viral load (4.46 ± 0.66 vs 4.90 ± 0.81, p=0.020) than those without stop codons in the vif ORF. Interstingly, the G→A substitution rate of participants with defective *vif* (2.73 ± 0.38 100^{-1} bp) reflects that obtained *in vitro* utilising wild type 3G and Δ*vif* HIV-1 virions (2.98 100^{-1} bp)[6]

CONCLUSIONS

• The G →A load is bimodally distributed around mean values of 4% and 12% of G content.

• At least 40% of the sampled cohort had evidence of APOBEC3G-mediated G→A mutations with a TGG context preference.

• APOBEC3G, rather than APOBEC3F, was the dominant contributor to G→A hypermutation.

• Participants harbouring proviral DNA with >10.6% G→A load (20% of sampled cohort) had significantly reduced viral load compared to those with proviral DNA with <10.6% G→A load.

•The influence of G→A hypermutation on viral load was independent of that exerted by HLA-B57, -B58 and/or CCR5Δ32 alleles.

•*Vif* in-frame stop codons and 3 APOBEC3G non-coding SNPs were associated with hypermutation.

REFERENCES

[1] Bishop *et al. Science.* 2004;305(5684):645. [2] Kao *et al. Retrovirol.* 2004;1(1):27. [3] Liddament *et al. Curr Biol.* 2004;14(15):1385-91. [4] Wiegand *et al. EMBO.* 2004;23(12):2451-8. [5] Bishop *et al. Curr Biol.* 2004;14(15):1392-6. [6] Newman *et al, Curr. Biol.* 15, 2 (2005).